INVESTIGATION OF SELECTED COLLAGEN GENES IN EXERCISE-RELATED MUSCULOSKELETAL SOFT TISSUE PHENOTYPES.

By

Kevin O'Connell

This thesis is presented for the degree of

DOCTOR OF PHILOSOPHY

In the Department of Human Biology

Faculty of Health Sciences

University of Cape Town

South Africa

August 2014

Supervisors

Prof Malcolm Collins and Dr Michael Posthumus

UCT/MRC Research Unit for Exercise Science and Sports Medicine

Sports Science Institute of South Africa

Boundary Road, Newlands, 7700

Cape Town

South Africa

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Thank you to my supervisors, Prof Malcolm Collins and Dr Mike Posthumus. The completion of this thesis would literally not have been possible without your guidance and support. Your patience and understandings kept this thesis on track when things got dodgy.

Thanks to all the staff and students at ESSM that make it such a great place to work. Thanks to Stef for keeping things fun with sessions of Stickman Cricket and Broforce. To my office mates Ben, James, Colleen and Julia, thanks for keeping me hard-working and laughing every day. Thanks to Neezaam and Trevino for all their help in the lab, and to Fiona for always sorting out the funding payouts.

Thanks to all the participants who donated time and DNA to these studies. Also thanks to all the people that were responsible for recruiting them.

Thanks to my girlfriend, Bridget. Your love and support have kept me going through all of my studies.

Thanks to my mom and dad for all your continued support of all of my studies. Thank you for your generosity in funding my studies and for supporting me when things got lean. Thank you for all the opportunities that you have given me, without you I would never have gotten to where I am.

This thesis was supported, in part, by funds from the National Research Foundation (NRF) of South Africa, the University of Cape Town (UCT) and the Medical Research Council (MRC) of South Africa.

LIST OF SCIENTIFIC OUTPUTS FROM THIS THESIS

ARTICLES IN INTERNATIONAL PEER-REVIEWED JOURNALS

1) **O'Connell K**, Posthumus M, Collins M. *COL6A1* gene and Ironman triathlon performance. *International Journal of Sports Medicine.* 2011;32(11):896-901.

2) **O'Connell K**, Saunders CJ, Collins M. Collagen Gene Sequence Variants in Exercise-Related Traits. *Central European Journal of Sports Science and Medicine* 2013; 1: 3-17

3) **O'Connell K**, Posthumus M, Collins M. No association between *COL3A1*, *COL6A1* or *COL12A1* gene variants and range of motion. *Journal of Sports Sciences.* 2013;31(2):181-7.

4) **O'Connell K**, Posthumus M, Schwellnus MP, Collins M. Collagen genes and exercise-associated muscle cramping. *Clinical Journal of Sport Medicine*. 2013;23(1):64-9.

5) **O'Connell K**, Posthumus M and Collins M. Collagen gene interactions and endurance running performance. *South African Journal of Sports Medicine*. 2014;26(1):9-14

6) **O'Connell K**, Knight H, Ficek K, Leonska-Duniec A, Maciejewska-Karlowska A, Sawczuk M, Stepien-Slodkowska M, O'Cuinneagain D, van der Merwe W, Posthumus M, Cieszczyk P, Collins M. Interactions between Collagen Gene Variants and risk of Anterior Cruciate Ligament Rupture. *European Journal of Sport Science*. In Press.

PRESENTATIONS AT INTERNATIONAL CONGRESSES

1) **O'Connell K**, Posthumus M, Collins M. The *COL3A1*, *COL5A1* and *COL6A1* genes and exercise associated muscle cramping (EAMC). European College of Sports Science, Liverpool, England (2011). (Oral presentation, presented by M Posthumus).

2) **O'Connell K**, Posthumus M, Collins M. The *COL6A1* gene and performance in the South African Ironman triathlon. European College of Sports Science, Liverpool, England (2011). (Oral presentation, presented by M Posthumus).

3) **O'Connell K**, Posthumus M, Collins M. The *COL6A1* gene and performance in the South African Ironman triathlon. Joint International Conference of the African and Southern African Societies of Human Genetics, CTICC, Cape Town, South Africa (2011). (Poster presentation).

4) **O'Connell K**, Posthumus M, Collins M. Investigation of *COL3A1*, *COL6A1* and *COL12A1* polymorphisms with Anterior Cruciate Ligament Rupture in

Two Independent Populations. European College of Sports Science, Barcelona, Spain (2013). (Oral presentation).

PRESENTATIONS AT LOCAL CONGRESSES

1) **O'Connell K**, Posthumus M, Schwellnus M, Collins M. The *COL5A1* gene and Exercise Associated Muscle Cramping (EAMC). MRC Research Day, MRC, Cape Town, South Africa (2010). (Oral presentation).

2) **O'Connell K**, Posthumus M, Collins M. The *COL6A1* gene and performance in the South African Ironman triathlon**.** Young Researchers Forum, University of Cape Town, South Africa (2011). (Oral presentation).

3) **O'Connell K**, Posthumus M, Collins M. Collagen genes and exercise associated muscle cramping in Ironman triathletes**.** Experimental Biology Group Seminar, University of Stellenbosch, South Africa (2011). (Oral presentation).

4) **O'Connell K**, Ficek K, Leonska-Duniec A, Maciejewska-Karlowska A, Sawczuk M, Stepien-Slodkowska M, O'Cuinneagain D, van der Merwe W, Posthumus M, Cieszczyk P, Collins M. Collagen gene variants and risk of anterior cruciate ligament rupture in two independent populations. Clinical laboratory sciences and human biology research day, University of Cape Town, South Africa (2013). (Poster presentation)

ABBREVIATIONS

ACL:	Anterior cruciate ligament
ANOVA:	One-way analysis of variance
AUS:	Australia
BMI:	Body mass index
bp:	Base pair
CI:	Confidence interval
CoB:	Country of Birth
COL1A1:	Gene encoding the α1 chain of type I collagen
COL3A1:	Gene encoding the α1 chain of type III collagen
COL5A1:	Gene encoding the α1 chain of type V collagen
COL6A1:	Gene encoding the α1 chain of type VI collagen
COL12A1:	Gene encoding the α1 chain of type XII collagen
CON:	Control participants.
CoR:	Country of Residence
DISH:	Diffuse idiopathic skeletal hyperostosis
DMEM:	Dulbecco's modified Eagle's medium
DNA:	Deoxyribonucleic acid

EAMC: Exercise associated muscle cramps

ECM: Extracellular matrix

EDS: Ehlers-Danlos syndrome

EDTA: Ethylenediaminetetraacetic acid

FACIT: Fibril associated collagen with interrupted triple helices

FCS: Fetal calf serum

HWE: Hardy-Weinberg Equilibrium

Kb: Kilo base pairs

mRNA: Messenger ribonucleic acid

NON: Participants with a non-contact mechanism of ACL rupture (Chapter 3)

NON: Participants with no life-long history of EAMC (Chapter 7)

OR: Odds ratio

OPLL: Ossification of the posterior longitudinal ligament

p: Short arm of a chromosome

PCR: Polymerase chain reaction

PBS:	Phosphate buffered saline
q:	Long arm of a chromosome
qRT-PCR:	Quantitative real-time polymerase chain reaction
RFLP:	Restriction fragment length polymorphism
ROM:	Range of Motion
SA:	South African
SEE:	Standard error of the estimate
ShTR:	Total shoulder rotation
SLR:	Straight leg raise
SNP(s):	Single nucleotide polymorphism(s)
TEN:	Tendinopathy participants
UTR:	Untranslated region

LIST OF FIGURES FROM THIS THESIS

LIST OF TABLES FROM THIS THESIS

ABSTRACT

INTRODUCTION

Previous findings have suggested that functional variants within collagen encoding genes are associated with several musculoskeletal soft tissue injuries and other exercise-related phenotypes. Specifically variants within the functional *COL5A1* 3'-untranslated region (UTR) have previously been associated with (1) chronic Achilles tendinopathy, (2) Anterior Cruciate Ligament (ACL) ruptures in females, (3) endurance running performance and (4) range of motion (ROM). Since this gene encodes for an important structural component of the collagen fibril it has been hypothesised that variants within other collagen fibril encoding genes, such as *COL3A1*, *COL6A1* and *COL12A1*, will also be associated with these and/or other musculoskeletal soft tissue injuries and exercise-related phenotypes.

The *COL5A1* rs12722 and *COL12A1* rs970547 gene variants have been previously associated with risk of ACL ruptures in females [153;154] and/or chronic Achilles tendinopathy [131;181]. The first aim of this thesis was therefore to investigate the *COL3A1* rs1800255 and *COL6A1* rs35796750 gene variants as risk factors for these musculoskeletal soft tissue injuries. The objectives to address this aim were:

- To investigate the *COL3A1* rs1800255 and *COL6A1* rs35796750 gene variants as risk factors for ACL ruptures in a South African cohort.

- To investigate the *COL6A1* rs35796750 gene variant as a risk factor for Achilles tendinopathy in independent SA and Australian cohorts.

Since the *COL5A1* rs12722 variant was previously associated with endurance performance and ROM, the second aim of this thesis was to further investigate the *COL3A1*, *COL6A1*, and *COL12A1* genes for associations with these exercise-related phenotypes. Therefore, the objectives to address this aim were:

- To investigate the *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 gene variants as intrinsic factors for ROM in a SA cohort.
- To investigate the *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 gene variants as intrinsic factors for endurance performance in participants of the SA Ironman triathlon.

The third aim of this thesis was the investigation of *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 in novel exercise-related phenotypes, namely exercise associated muscle cramps (EAMC) and rugby union playing level and position. The objectives to address this aim were:

- To investigate the *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 gene variants as risk factors for EAMC in a South African cohort.
- To investigate the *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 gene variants as intrinsic factors for

rugby union playing level and position in professional SA Super 15 rugby union players.

Since *COL6A1* rs35796750 has previously been predicted to be functional, the final aim of this thesis was to investigate the expression of the *COL6A1* gene in individuals with known rs35796750 genotypes.

METHODS

Participants within several phenotypic groups were previously recruited. These groups include; (1) participants with diagnosed ACL rupture and matched controls (2) participants with diagnosed chronic Achilles tendinopathy and matched controls, (3) participants that were measured for range of motion, (4) participants from the South African Ironman triathlon, (5) participants with a history of EAMC and matched controls and (6) professional South African Super 15 rugby union players. All participants were genotyped, where appropriate, for *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 using standard PCR based methods. Furthermore, to address the final aim of this thesis primary skin fibroblasts were cultured, from participants with known *COL6A1* rs35796750 genotypes, and used to determine *COL6A1* gene expression.

RESULTS

COL5A1 rs12722

This thesis identified a novel independent association between *COL5A1* rs12722 and past history of EAMC but not for rugby union playing level and position. Specifically the *COL5A1* rs12722 CC genotype was significantly under-represented (p=0.031, OR=2.2, 95% CI 1.1–4.5) in the EAMC group (CC genotype 11.1%) when compared to the control group (CC genotype 21.8%).

COL12A1 rs970547

Although this *COL12A1* variant was previously associated with ACL ruptures in females, no additional significant independent associations were identified between *COL12A1* rs970547 and any of the exercise-related phenotypes (ROM, EAMC, endurance performance and rugby union) investigated in this thesis.

COL6A1 rs35796750

An additional novel finding of this thesis was that the *COL6A1* rs35796750 gene variant was independently associated with endurance cycling performance in Ironman triathletes. Specifically, participants with the *COL6A1* rs35796750 TT genotype were significantly faster than those with a TC or CC genotype in the bike component of the triathlon (p=0.014). No other independent associations were identified between this variant and any of the other musculoskeletal soft tissue injuries and exercise-related phenotypes investigated in this thesis.

Novel functional analysis of the *COL6A1* rs35796750 variant revealed, for the first time, that participants with a *COL6A1* rs35796750 T allele (TT or TC genotype) had significantly higher levels of *COL6A1* gene expression than their CC genotype counterparts. Specifically, participants with a T allele had a 1.8 fold increase in *COL6A1* expression (p=0.001) when compared to participants with a CC genotype.

COL3A1 rs1800255

The *COL3A1* rs1800255 gene variant was significantly independently associated with the male forwards rugby union positional sub-group. Specifically, the *COL3A1* rs1800255 GG genotype was significantly (p=0.047) under-represented in the forwards rugby union position sub-group (39.6%, n=38) when compared to controls (53.5%, n=76). No other independent associations were identified between this variant and any of the other musculoskeletal soft tissue injuries and exercise-related phenotypes investigated in this thesis.

In addition to the independent associations identified, a final novel finding of this thesis was that a number of inferred pseudo-haplotypes, constructed from *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and/or *COL12A1* rs970547, were associated with the exercise-related phenotypes investigated in this thesis.

CONCLUSIONS

This thesis identifies a number of collagen gene variants which are associated with various musculoskeletal soft tissue injuries and other exercise-related phenotypes. Interestingly, the functional *COL6A1* rs35796750 variant was associated with most of

the phenotypes investigated in this thesis. Furthermore, as initially hypothesised, the mechanism for most of these associations is consistent with that previously proposed for *COL5A1* rs12722 [41]. Future studies are required to investigate and understand the contrasting and unique results identified for *COL6A1* rs35796750. The *COL3A1* rs1800255 variant was associated with a number of phenotypes, while no additional independent associations were identified for *COL12A1* rs970547. In addition, varying gene-gene interactions were identified between the four variants and six phenotypes investigated in this thesis. These results highlight the complexity in identifying genetic, and other, markers for the inter-individual predisposition to musculoskeletal soft tissue injuries and other exercise-related phenotypes.

CHAPTER 1

INTRODUCTION AND SCOPE OF THE THESIS

Collagen proteins are important structural components of the collagen fibril and building blocks of musculoskeletal tissues. Rare mutations within collagen encoding genes, such as *COL1A1*, *COL3A1*, *COL5A1*, *COL6A1* and *COL12A1*, cause debilitating musculoskeletal disorders. These disorders include but are not limited to, osteogenesis imperfecta [144;147], chondrodysplasias [205], Ehlers-Danlos syndrome [34;121;122;124;186;218], Bethlem Myopathy [17;161] and Ullrich Congenital Muscular Dystrophy [17;161]. These severe clinical disorders highlight the importance of this protein family in the normal structure and function of musculoskeletal tissues and suggest a lack of biological redundancy within the collagen fibril [41]. Musculoskeletal tissues, are not only common sites of injury during participation in physical and certain occupational activities, they also play an important role in contributing to athletic performance. Many of these collagen fibril encoding genes are therefore ideal candidates for association with less severe multifactorial exercise-related phenotypes [163].

Changes to types III, V, VI and XII collagen levels within the collagen fibril result in changes to its architecture and organisation which may modulate the biomechanical properties of tissue containing these proteins [19;21;78;110;133;207]. It has therefore been proposed that common variants, like the previously studied *COL5A1*

rs12722 [1;2;28;29;40;107;131;151;153;156;179] and *COL12A1* rs970547 [154;181] single nucleotide polymorphisms, within the genes that encode these collagens may be associated with seemingly unrelated musculoskeletal soft tissue injuries and other exercise-related phenotypes, such as range of motion and athletic performance [41]. The *COL5A1* and *COL12A1* genes encode for the α1(V) and α1(XII) chains of types V and XII collagen respectively [18]. It has been previously proposed that the associated collagen gene variants will, at least in part, directly or indirectly alter collagen fibril architecture and structure and, thereby, the mechanical properties of musculoskeletal soft tissue [41]. This thesis will therefore examine the hypothesis that common variants within several collagen-encoding genes are associated with musculoskeletal soft tissue injuries and other exercise-related phenotypes.

The *COL5A1* rs12722 variant was previously shown to be associated with chronic Achilles tendinopathy [131;179] and anterior cruciate ligament (ACL) ruptures in females [153], while the *COL12A1* rs970547 variant was only associated with ACL ruptures in females [154], but not Achilles tendinopathy [181]. The first aim of this thesis was therefore to further investigate the association of common variants within the *COL3A1* and *COL6A1* genes as additional risk factors for chronic Achilles tendinopathy and ACL ruptures. These genes encode for the α1 chains of types III and VI collagen respectively. In addition, the *COL5A1* rs12722 variant was also previously shown to be associated with range of motion [1;28;29;40] and athletic performance [1;28;151]. The second aim of this thesis was therefore to investigate the *COL3A1, COL6A1* and *COL12A1* genes and these exercise-related phenotypes. The third aim of this thesis was to investigate the collagen genes, specifically *COL3A1*, *COL5A1*, *COL6A1* and *COL12A1*, for associations with novel exercise-

related phenotypes, namely exercise associated muscle cramps and rugby union playing level and position. The final aim of this thesis was to investigate the functional effects of the *COL6A1* rs35796750 variant in order to further the understanding, and determine biological mechanisms, of how this gene variant may modulate exercise-related phenotypes.

In preparation for the exploration and further discussion of the experimental chapters of this thesis, Chapter 2 will provide a focussed review of collagens, their genes, as well as the role of collagen gene sequence variants in the aetiology of exercise-related phenotypes. The rationale for the collagen gene sequence variants investigated in this thesis will also be provided. Subsequent experimental chapters (Chapters 3 to 8) will use a candidate gene approach to achieve the primary, secondary and tertiary aims of this thesis. The final aim of this thesis will be achieved through the use of specific molecular assays and techniques as outlined in chapter 9. An overall conclusion and discussion will be presented in the final chapter of this thesis.

CHAPTER 2

COLLAGEN GENE SEQUENCE VARIANTS AND EXERCISE-RELATED PHENOTYPES: A REVIEW

The information presented in this chapter has been published in a condensed form in the peer-reviewed article: O'Connell K, Saunders CJ, Collins M. Collagen Gene Sequence Variants in Exercise-Related Traits. *Central European Journal of Sports Science and Medicine* 2013; 1: 3-17

2.1 AN OVERVIEW OF COLLAGENS AND THE STRUCTURE OF THE COLLAGEN FIBRIL

Collagens are a family of twenty-eight structurally and functionally diverse proteins which consist of three polypeptide α-chains wound in a characteristic uninterrupted or interrupted left-handed triple helical structure [84;114]. The majority of these proteins are important for tissue assembly or maintenance [84]. Furthermore, many are the major macromolecular building blocks of the collagen fibril, which is the structural component of tendons, ligaments, bone, cartilage and connective tissue structures in muscle and other tissues (Figure 2.1) (Section 2.5.1). Although some collagen types, such as type XI collagen [16], are structural components or associated with the collagen fibril in both types of tissues, non-cartilaginous connective tissues generally contain a different set of collagen proteins when compared to cartilage (Table 2.1) (Figure 2.1) [9;84]. Since the collagen fibril within

the musculoskeletal system is an important structure in injury and exercise-related phenotypes, this review will focus predominately on the structural collagens of the non-cartilage and, to a lesser extent, the cartilage fibril as well as those collagens that are directly associated with these fibrils. The structure and functions of the entire collagen family has been previously reviewed [84;164].

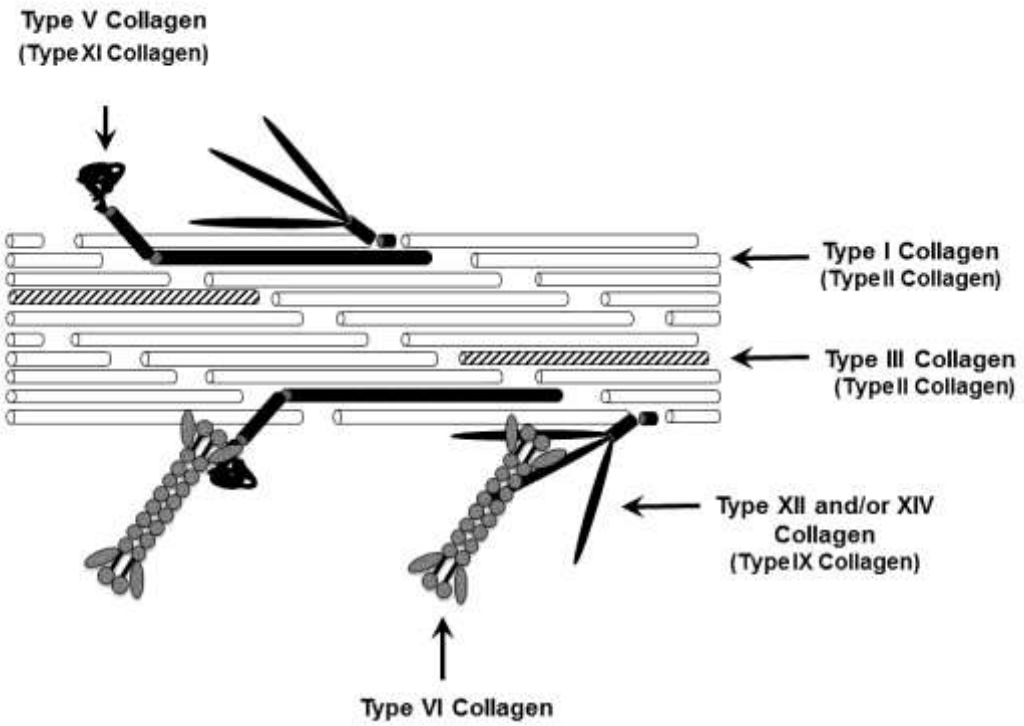


Figure 2.1. A schematic diagram of the collagen fibril, adapted from Collins and Posthumus [41]. The major fibrillar type I (solid white cylinders) and III (hatched cylinders) collagen molecules comprise the majority of the non-cartilaginous fibril. The minor fibrillar type V collagen (solid black cylinders with globular ends) is embedded within the fibril with a protruding amino terminal end. Type XII and XIV collagen (solid black cylinders with forked ends) belong to the fibril-associated collagens with interrupted triple helices (FACITs) sub-family. Beaded non-fibrillar type VI collagen (solid grey circles) interacts, at the surface of the fibril, with the other collagens that make up the fibril. Collagen types listed in parentheses are the cartilage specific collagens equivalent to those they are listed under. The proteins in this diagram are not drawn to scale.

Table 2.1. The classification of collagens within the non-cartilage and cartilage fibril, as well as, the basement membrane (BM) area.

		Non-cartilage	Cartilage	BM Area
Fibrillar	Major	I, III	II	III [2]
	Minor	V	XI [1]	
	"Non-classical"	XXIV [3]	XXVII [3]	
Non-fibrillar	FACITs [4]	XII, XIV, *XVI* [5], XX [6], XXI [7]	IX, XX [6]	XIX [7], XXII [8]
	Beaded filament			VI [9], *XXVIII* [10]
	Network forming			IV
	Short chain	*VIII*	*X*	*VIII*, *X*
	Anchoring fibrils			*VII* [11]
	Multiplexins [12]			XV [13], *XVIII*
	Transmembrane domains			XIII, *XVII*, XXIII, XXV
	Unknown function	*XXVI* [14]		

Collagens listed in italic typeset and underlined are not, or have not been shown to be, involved in the musculoskeletal system.

[1] also expressed in the developing tendon

[2] reticular fibres

[3] not classical fibrillar collagens, expressed primarily in cartilage and sites of transition from cartilage to bone

[4] fibril associated collagens with interrupted triple helices

[5] high levels have been detected in fibroblasts, keratinocytes and smooth muscle [89]

[6] prevalent within corneal epithelium but has also been identified in embryonic skin and tendons [97]

[7] been identified within skeletal muscle amongst other tissues [54;132]

[8] expressed in myotendinous junctions [99]

[9] predominately expressed in the basement membranes of skeletal muscle [109]

[10] expressed in the skin and calvaria [202].

[11] expressed only in the retina [149]

[12] contains multiple triple-helix domains interrupted by non-collagenous domains

[13] mainly derived from muscle cells and fibroblasts [68]

[14] specifically expressed in the ovary and testis [169]

The collagen family can be divided into two groups based on their structure and function: (i) the fibrillar, or fibril-forming, collagens which form the fibrillar scaffolding for the extracellular matrix and, (ii) the non-fibrillar collagens which include, amongst others, the fibril associated collagens with interrupted triple helices (FACITs), beaded filament collagens, network forming collagens, short chain collagens, anchoring fibrils, multiplexin and transmembrane domains (Table 2.1) [84;166]. The different musculoskeletal collagen types, with a focus on collagens found within tendons, ligaments and/or connective tissue layers of skeletal muscle, will be reviewed in the following sections.

2.1.1 Classical Fibrillar Collagens

The classical fibrillar collagens are further sub-divided, based on quantity, into the major (types I, II, III) and minor (types V, XI) fibrillar collagens [84]. The predominant major fibrillar collagen is type I collagen consisting of two α1(I) and one α2(I) chains. It is responsible for the hierarchical structure and mechanical strength of non-cartilage connective tissue [52;75;162]. Type II collagen, a homotrimer consisting of three α1(II) chains, is the equivalent main structural collagen in cartilage [86]. Type III collagen, which consists of three α1(III) chains, is also a major fibrillar collagen which forms heterotypic fibrils together with type I collagen, and is important in healing and during fibrillogenesis [9;114]. It is thought that type III collagen regulates the diameter of type I collagen fibrils during development and healing by limiting lateral growth [9]. Tissues with elastic properties, such as skin and arteries, also have a higher type III collagen content than less elastic tissues [60]. Type III collagen

is also known to form a fine supporting meshwork known as the reticular fibres within the basement membranes of soft tissues [200].

Type V collagen is a minor fibrillar collagen which is co-expressed with types I and III collagen and also plays a role in the regulation of type I collagen fibril diameter [18;201]. The predominant isoform of type V collagen is a heterotrimer of two α1(V) and one α2(V) chains, however other isoforms, which may contain α3(V) chains, also exist [18]. Type XI collagen is traditionally considered the minor fibrillar collagen co-expressed with type II collagen in cartilage [84;201]. Recent investigations on the function of type V collagen in tendon development suggest a synergistic role with type XI collagen in the regulation of fibrillogenesis and growth of mature tendon fibrils [207]. Types XXIV and XXVII collagen, although not classical fibrillar collagens, are nevertheless also fibrillar collagens expressed primarily in cartilage and sites of transition from cartilage to bone [23;74;98;144], while type XXVII collagen is also expressed in skeletal muscle [23;74;144].

2.1.2 Fibril-Associated Collagen with Interrupted Triple Helices (FACITs)

The “classical” well characterised FACITs, which include types IX, XII and XIV [60], are important non-fibrillar collagens in mediating cell-matrix interactions between the collagen fibres and cell surfaces [166], and during stabilisation of the attached collagen triple helices [25;164]. Figure 2.2 shows the structure of these “classical” FACITs, as well as how they affect cell-matrix interactions. Type IX collagen plays a role in the mechanical stability and resistance to swelling of the collagen type II fibril framework of cartilage [209]. One of the short isoforms of type XII collagen, XIIB-1, is predominately expressed in tendons and ligaments [85]. Type XIV collagen is also shown to be expressed in skeletal muscle connective tissue after denervation and during muscle fibre regeneration [195].

Other lesser characterised FACITs include types XVI, XX, XXI, XIX and XXII collagens. High levels of type XVI collagen have been detected in fibroblasts and keratinocytes, as well as smooth muscle [89]. Type XX collagen is prevalent within corneal epithelium but has also been identified in embryonic skin and tendons [97]. Types XIX and XXI collagen have been identified within skeletal muscle amongst other tissues [54;132], while type XIX collagen is also expressed in the skin [132]. Furthermore, investigation of the myotendinous junctions has revealed that muscle cells produce collagen XXII, which acts as a cell adhesion ligand for skin epithelial cells and fibroblasts [99].

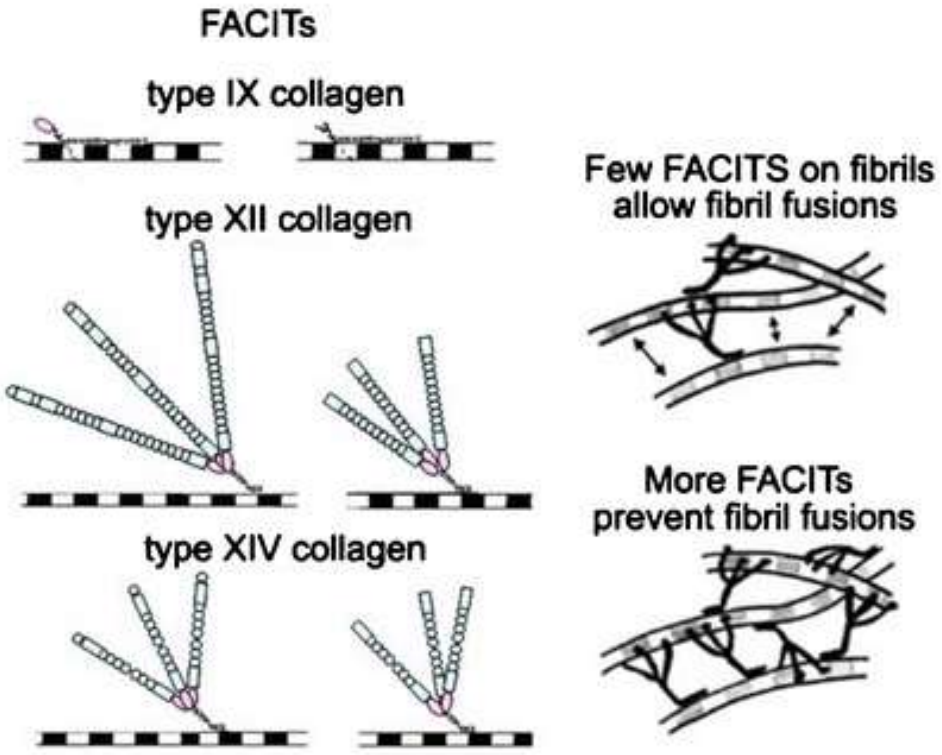


Figure 2.2. A schematic diagram showing the structures of types IX, XII and XIV collagen, adapted from Gordon and Hahn [64]. The cell-matrix interaction of FACITs is also shown. The FACITs are required for fibril fusion and interaction, however too many FACITs will adversely affect fibril fusion and interaction.

2.1.3 Non-fibrillar Beaded Filament Collagens

Type VI collagen (Figure 2.3), a non-fibrillar beaded filament collagen, is predominately expressed as a microfibrilllar network in close association with the basement membranes of skeletal muscle [109], and also facilitates fibrillar interaction with the extracellular matrix through interactions with types I, V, and XII collagens [129;190;201]. In addition, abnormal collagen fibrillogenesis in *Col6a1* deficient mice highlights the potential importance that type VI collagen may have in this biological

process [78]. Type VI collagen is predominately a heterotrimer made up of one α1(VI), one α2(VI) and one α3(VI) chain, encoded by the COL6A1, COL6A2 and COL6A3 genes respectively. More recently, heterotrimers containing α4(VI), α5(VI) and α6(VI) chains, encoded by the *COL6A4*, *COL6A5* and *COL6A6* genes, have also been identified [55]. Type XXVIII collagen is a beaded filament collagen restricted in expression to the skin and calvaria (skullcap), specifically to the dorsal root ganglia and peripheral nerves [202].

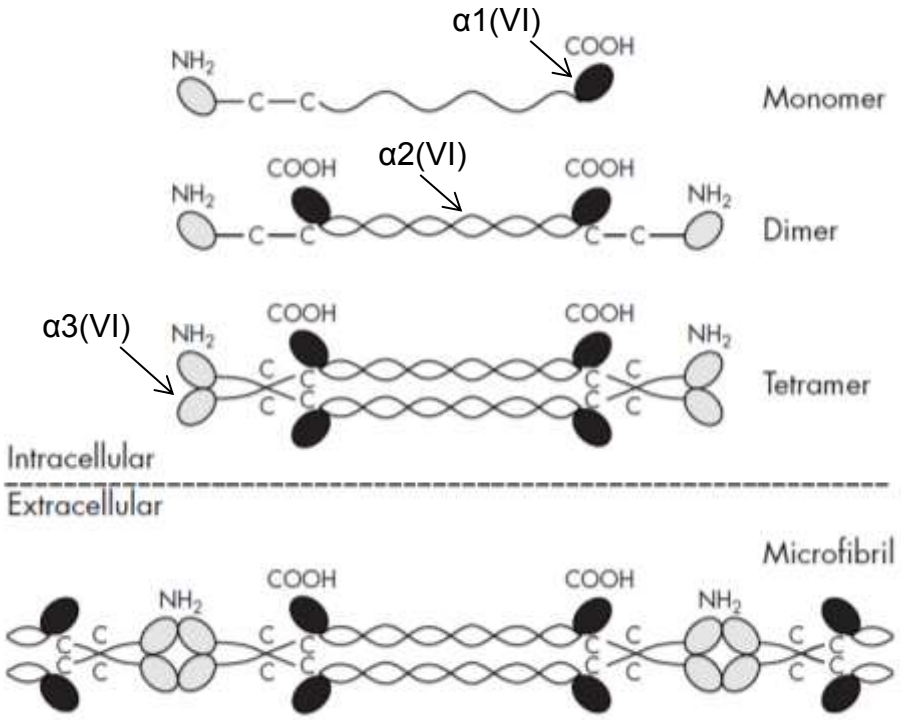


Figure 2.3. A schematic diagram of the type VI collagen beaded filament, adapted from Lampe and Bushby [109]. The α1(VI), α2(VI) and α3(VI) chains assemble intracellularly and the mature type VI collagen heterotrimer is then expressed extracellularly.

2.1.4 Other Non-fibrillar Collagens

Other non-fibrillar collagens include the network forming (IV), short chain (XVIII, X), anchoring fibrils (VII), multiplexins (XV, XVIII), transmembrane domains (XIII, XVII, XXIII, XXV) and those with unknown function (XXVI). Although type IV collagen isn't a structural component of the fibril, it does form a major part of the basement membrane, which surrounds and anchors the cellular component of musculoskeletal tissues, and maintains tissue architecture during development and wound healing [84;194]. In particular, the basement membranes are connected to cells by receptors which specifically bind to type IV collagen and other proteins.

The short-chain collagen types VIII and X form networks in the basement membranes of non-cartilaginous and cartilaginous tissues respectively [84]. Type VII collagen, the only anchoring fibril, is expressed only in the retina [149]. Multiplexins, types XV and XVIII collagen, contain multiple triple-helix domains interrupted by non-collagenous domains. Type XV collagen is mainly derived from muscle cells and fibroblasts [68], and the absence of type XV collagen results in skeletal myopathy and cardiovascular defects in mice [49]. Less is known about type XVIII collagen, however lack of type XVIII collagen leads to ocular defects as a result of aberrant blood vessel formation and integrity [211].

Although all transmembrane domain collagens, types XIII, XXIII and XXV share structural similarity, all differing to type XVII collagen. Type XIII collagen has high expression during development and growth which decreases during adulthood and is expressed in the basement membrane of myotendinous and neuromuscular

junctions, amongst other tissues [72;189]. Type XXV is also necessary for the normal functioning of the neuromuscular junction [192]. Type XXIII collagen shows a greater diversity of expression and has been identified in the basement membrane of lung, cornea, skin, tendon and kidney tissue [100]. Unlike types XIII, XXII and XXV collagen, type XVII collagen is not implicated in the musculoskeletal system but is instead necessary for the normal supply of melanocytes to the epidermis [65]. The non-fibrillar collagens are further reviewed by Ricard-Blum et al. [164].

2.1.5 Collagen Synthesis

Typically, the polypeptide α-chains of the fibrillar collagens comprise approximately 1000 amino acids in an uninterrupted repeating triplet, which contains a glycine amino acid in every third position with proline and 4-hydroxyproline residues in the other two positions (Gly-X-Y) [84]. These repeating triplets are essential for the formation of the uninterrupted and interrupted triple helical domains in all the collagens (Figure 2.4) [84]. Most collagens occur as homotrimers of three identical α-chains, however heterotrimers of two or three different α-chains are also common [84]. These α-chains are wound together in a left-handed triple helix and, in the major fibrillar collagens (I, II and III), the globular amino- and carboxy-terminal domains flanking the uninterrupted triple helix are post-translationally cleaved prior to their aggregation into collagen fibrils (Figure 2.4) [9;84;136]. Only the carboxy-terminal globular domains of the minor fibrillar collagens (V and XI) are post-transnationally cleaved. The triple helical domains of types V and XI collagens are imbedded in the non-cartilaginous and the cartilaginous fibrils respectively. The amino-terminal globular domains of these molecules protrude from the surface of the

fibril where they play an important role in regulating fibrillogenesis (Figure 2.1). The synthesis of the collagen proteins has been extensively reviewed by Banos et al. [9].

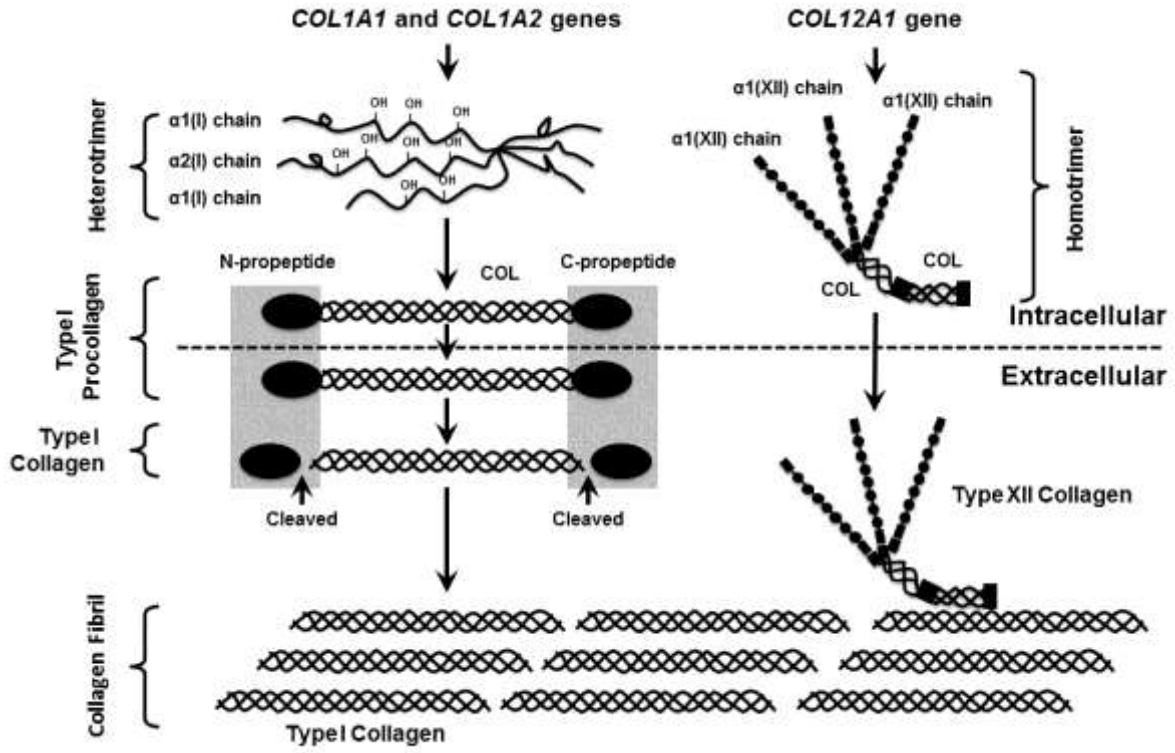


Figure 2.4. A schematic diagram of collagen fibril assembly. The polypeptide α-chains of the fibrillar collagens comprise an uninterrupted repeating triplet, which contains a glycine amino acid in every third position with proline and 4-hydroxyproline residues in the other two positions (Gly-X-Y). This is indicated by the hydroxyl (-OH) groups on the type I collagen α-chains. These repeating triplets form the triple helical domains in all the collagens. In the major fibrillar collagens, including type I collagen, the globular carboxy- and amino-terminal domains flanking the triple helix are post-translationally cleaved prior to their aggregation into collagen fibrils. Post-translational cleavage of the amino-terminal domain does not occur in some of the minor fibrillar (type V collagen) and non-fibrillar (type XII and XIV collagens) collagens. Type I and type XII collagen have been used as examples for the collagen fibril assembly process. The molecules and proteins in this diagram are not drawn to scale.

Although several collagen types are expressed in tendons, ligaments and the connective tissue structures within skeletal muscle, this thesis will focus on types I, III, V, VI and XII collagen. These collagens will be further reviewed in section 2.3 and 2.6 of this chapter, as well as, the introductions of the results chapters of this thesis.

2.2 MENDELIAN AND MULTIFACTORIAL COLLAGEN RELATED DISORDERS

The major structural domains of collagen proteins are typically highly conserved across species, and each polypeptide α-chain is encoded by a specific gene (Table 2.2). In the case of homotrimeric collagens, all three α-chains are encoded by the same gene. For example, the α1 chains of type II collagen are encoded by the *COL2A1* gene. In heterotrimeric collagens, each different α-chain is encoded by a different gene. The *COL1A1* and *COL1A2* genes, for example, encode the α1 and α2 chains of type I collagen respectively. Rare mutations in many of these genes result in serious musculoskeletal disorders including, amongst others, osteogenesis imperfecta [144;147], chondrodysplasias [205] and Ehlers-Danlos syndrome [34;121;122;124;186;218] (Table 2.2). Symptoms of these disorders include bone fragility, joint hypermobility, skin hyperextensability and abnormal wound healing [124;147]. These severe clinical disorders highlight the importance of this protein family in the normal structure and function of musculoskeletal tissues.

Table 2.2. Disorders resulting from rare mutations or common variants in the collagen gene(s) that encode for the non-cartilage and cartilage fibril, as well as the collagen network around the fibril producing cells.

Type	Gene(s)	Mendelian Disorder(s)[a]	Complex Multifactorial Disorder(s)
I	*COL1A1* *COL1A2*	Osteogenesis imperfecta, EDS, Infantile Cortical Hyperostosis	Osteoporotic Fractures, Osteoarthritis, Myocardial Infarction, Lumbar Disc Disease, Stress Urinary Incontinence
II	*COL2A1*	Collagenopathy, types II and III	Mild Spondyloepiphyseal Dysplasia Congenita, Nonsyndromic Cleft Palate, Osteoarthritis, Myopia
III	*COL3A1*	EDS (Hypermobility type; type 3), EDS (Vascular type; type 4)	Mitral Valve Prolapse, Pelvic Organ Prolapse, Cervical Artery Dissection, Rupture of Intracranial Aneurysms
IV	*COL4A1* *COL4A6*	Porencephaly, Alport Syndrome, Hematuria, Brain Small Vessel Disease	Arterial Stiffness
V	*COL5A1* *COL5A2* *COL5A3*	EDS (Classic type; types 1 and 2)	Central Corneal Thickness, Ischemic Heart Disease, Cervical Artery Dissection, Rupture of Intracranial Aneurysms
VI	*COL6A1* *COL6A2* *COL6A3*	Bethlem Myopathy, Ullrich Congenital Muscular Dystrophy	Ossification of the Posterior Longitudinal Ligament (OPLL) and the Ligamentum Flavum, Diffuse Idiopathic Skeletal Hyperostosis
VIII	*COL8A1* *COL8A2*	Posterior Polymorphous Corneal Dystrophy, Fuchs' Endothelial Corneal Dystrophy	Primary Open Angle Glaucoma, Central Corneal Thickness
IX	*COL9A1* *COL9A2* *COL9A3*	Epiphyseal Dysplasia, Stickler Syndrome	Osteoarthristis, Lumbar Disc Degeneration
X	*COL10A1*	Metaphyseal Chondrodysplasia	Macular Degeneration, Osteoarhtritis
XI	*COL11A1* *COL11A2* *COL2A1*	Fibrochondrogenesis, Marshall Syn., Stickler Syn., Weissenbacher-Zweymuller Syn., Deafness, Otospondylomegaepiphyseal Dysplasia	Lumbar Disc Herniation, Limbus Vertebra, Lumbar Spine Stenosis, OPLL, Rheumatoid Arthritis
XII	*COL12A1*	EDS, Developmental Delay, Mild Dysmorphism, Lax Connective Tissue	Abnormal Corneal Endothelial Maturation
XIV	*COL14A1*	Punctate Palmoplantar Keratoderma, Recurrent Strokes	Abnormal Corneal Endothelial Maturation
XXVII	*COL27A1*	None identified to date	Tourette's Syndrome

EDS, Ehlers-Danlos Syndrome; OPLL, Ossification of the Posterior Longitudinal Ligament.

[a] as listed in the OMIM database [142]

Since rare mutations within most of the genes encoding the collagens in the fibril cause severe disorders, it has been hypothesised that there is a lack of biological redundancy within the collagen fibril [41]. In support of this hypothesis, in addition to these severe rare Mendelian disorders, common variants within the collagen genes are also associated with a multitude of anomalous connective tissue phenotypes (Table 2.2). Therefore, it is possible that a continuum of associated phenotypes may exist for the range of genetic variation within these genes encoding the fibrillar and non-fibrillar collagens. An example of this genetic continuum was recently proposed to describe phenotypes associated with variation within the *COL5A1* gene (Figure 2.5) [41]. Two functional copies of the *COL5A1* gene are required for normal development, structure and function of the connective tissue. Mutations that inactivate both copies of *COL5A1* result in death *in utero* [206], while mutations, which inactivate a single copy of the gene (haploinsufficiency) will result in classical types of Ehlers-Danlos syndrome (EDS) [124]. At this end of the continuum the mutations in *COL5A1* result in death or EDS, which occur regardless of environmental exposure and independent of other non-genetic risk factors [41]. At the opposite end of the continuum, functional common variants within *COL5A1* contribute to more complex and less severe phenotypes, and do not result in disorders. These multifactorial phenotypes arise as a result of the interaction between genetic and non-genetic factors modifying physiological responses to environmental exposures [41].

Lethal mutations have also been described within other fibrillar collagen encoding genes, such as *COL1A1*, *COL1A2* and *COL2A1*, and severe musculoskeletal tissue disorders caused by non-lethal mutations within most of the genes encoding

components of the collagen fibril have also been reported (Table 2.2). As with the *COL5A1* genetic continuum, these mutations result in death or severe disorders which occur regardless of environmental exposure and independent of other non-genetic risk factors (Figure 2.5). At the opposite end of the continuum, it may be proposed that common variants within the same collagen genes contribute to less severe phenotypes, including exercise-related phenotypes, by modifying the effects of environmental exposures (Figure 2.5). This proposal is supported by experimental studies investigating common variants in genes encoding fibrillar and fibrillar-associated collagens with exercise-related phenotypes and will be expanded on in the following sections of this review chapter.

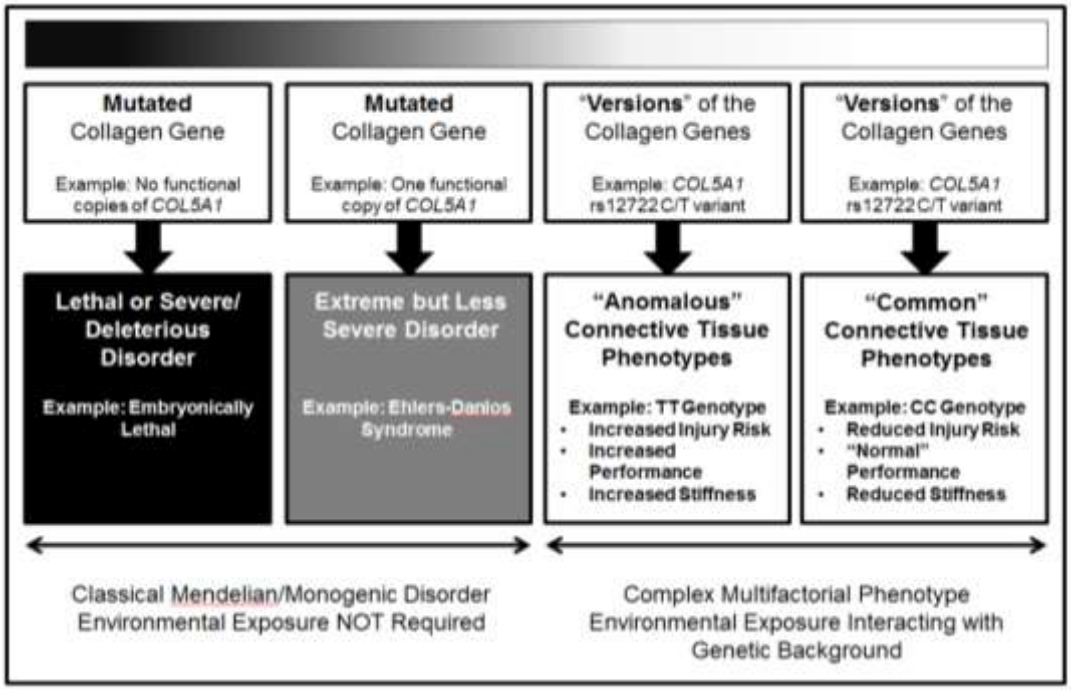


Figure 2.5. A proposed general genetic continuum for collagen genes, adapted from Collins and Posthumus [41]. The black shading represents the lethal or severe phenotypes due to mutations in the collagen genes. At this end of the continuum a single mutation results in the disorder. The white shading represents the most beneficial "versions" of the collagen genes. At this end of the continuum variants within collagen genes collectively contribute to the aetiology of the phenotype. The *COL5A1* gene and rs12722 gene variant are used as examples at each stage of the continuum.

2.3 FIBRILLAR AND FIBRIL-ASSOCIATED COLLAGEN GENES AND EXERCISE-ASSOCIATED PHENOTYPES

2.3.1 Type I Collagen

Type I collagen, encoded by the *COL1A1 and COL1A2* genes, is the predominant fibrillar collagen in non-cartilage connective tissues [52;75;162]. Several polymorphic regulatory elements have been identified within the promoter and first intron of the *COL1A1* gene, which result in differing transcription factor binding affinities and in turn regulate α1(I) chain production [81;126]. One of the most extensively investigated functional polymorphic transcription factor binding sites is the Sp-1 binding site variant (rs1800012, G/T) within the first intron of the *COL1A1* gene (Figure 2.6). The G allele of this variant produces normal levels of *COL1A1* mRNA and the heterotrimeric protein consisting of two α1(I) and one α2(I) chains. However, the T allele has been shown to functionally increase the binding of the transcription factor Sp-1 to its binding element within the first intron thereby increasing *COL1A1* mRNA transcription and translation, resulting in the proposed formation of α1(I) homotrimers from the excess α1(I) chains produced. It has been proposed that tissues consisting of both hetero- and homotrimeric type I collagen are produced, which have different mechanical properties to the normal tissue only consisting of heterotrimers [81;126].

The association of this functional Sp-1 binding site variant with cruciate ligament injuries has been investigated in Swedish [91], South African [152] and Polish [53] cohorts (Table 2.3). In all three studies the rare TT genotype was under-represented

in participants with diagnosed cruciate ligament (predominately ACL) ruptures when compared to apparently healthy control participants with no history of cruciate ligament injuries [53;91;152]. In a combined analysis of all three published studies and a second unpublished South African ACL cohort (H Knight, personal communication), which all had similar genotype distributions, the TT genotype was significantly over-represented ($p<0.0001$, odds ratio 23.8, 95% confidence interval 3.2 to 176.2) in the control groups (26 TT, 4.3%; 161 GT, 26.4% and 428 GG, 70.3%) when compared to the cruciate ligament rupture groups (1 TT, 0.2%; 168 GT, 31.4% and 366 GG, 68.4%) (Table 2.3) [42;53;91;152].

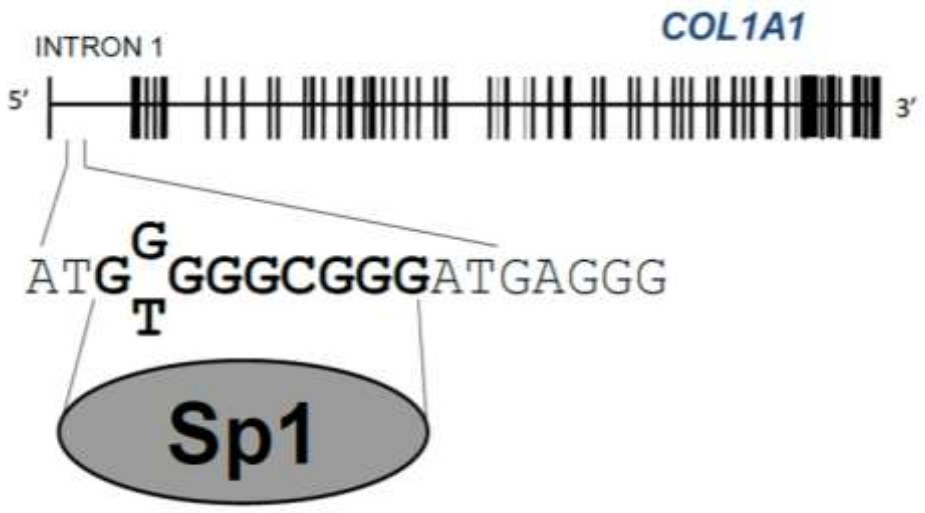


Figure 2.6. A schematic representation of the exon (vertical lines) and intron (horizontal lines) boundaries of the human *COL1A1* gene, which contains 52 exons. The Sp1 transcription factor binding site, which has been enlarged, is located within intron 1 and contains the rs1800012 G/T variant. The rs1800012 T allele increases the binding affinity of Sp1 thereby increasing COL1A1 mRNA and type I collagen protein translation. This variant is associated with cruciate ligament ruptures [53;91;152] and has been investigated as risk factor for other musculoskeletal soft tissue injuries [51;91;155].

Table 2.3. A comparison of the genotype frequency distributions for *COL1A1* rs1800012 (G/T) between anterior cruciate ligament (ACL) and control (CON) groups within the Swedish (SW), South African (SA) and Polish (PL) cohorts.

	***COL1A1* rs1800012 Genotype**			**n**	**Ref**
	GG	**GT**	**TT**		
SW CL [a]	69.5 (162)	30.0 (70)	0.5 (1)	233	[91]
SW CON	70.7 (230)	25.6 (83)	0.7 (12)	325	
SA ACL	65.9 (139)	34.1 (72)	0.0 (0)	211	[152][b]
SA CON	69.4 (102)	25.2 (37)	5.4 (8)	147	
PL ACL	71.4 (65)	28.6 (26)	0.0 (0)	91	[53]
PL CON	67.1 (96)	28.7 (41)	4.2 (6)	137	

Values are expressed as percentages with the number of participants indicated in parentheses. The total number (n) of participants genotyped in each individual or combined cohort is also indicated.

[a] This study did not distinguish between posterior and anterior cruciate ligament ruptures and merely grouped them together as cruciate ligament (CL) ruptures

[b] The published SA data was pooled with a second unpublished SA ACL cohort (H Knight, personal communication)

The most recent investigation of the Sp1 binding site variant (rs1800012) also reported an interaction between this variant and another polymorphic regulatory element in the *COL1A1* promoter (rs1107946, G/T, position −1997) in modulating risk of ACL injury [53]. Specifically, the haplotype constructed from the rs1800012 T and rs1107946 G alleles, both known to functionally increase expression of *COL1A1*, was over-represented in the control participants when compared to individuals with cruciate ligament ruptures [53] [81].

In the original Swedish study the intronic *COL1A1* rs1800012 TT genotype also showed a tendency to be under-represented (p=0.123) in participants with shoulder dislocations (0.8%) compared to controls (3.7%) (Table 2.4) [91]. In addition, the *COL1A1* Sp1 binding site variant was also investigated for risk of both acute Achilles tendon ruptures (n=41) and chronic Achilles tendinopathy (n=83) in a South African cohort (Table 2.4) [155]. Although no significant associations were identified for either the acute or chronic injuries, it is interesting to note that the TT genotype was absent in individuals diagnosed with acute Achilles tendon ruptures (0.0% ruptures, 2.4% tendinopathy and 4.8% controls) [155]. This variant was also recently shown not to be associated (p=0.170) with tennis elbow (lateral epicondylitis), although the TT genotype frequency was lower in the tennis elbow group (1.9%) when compared to the control group (6.8%) (Table 2.4) [51].

Furthermore, studies have shown an increase in type I collagen mRNA expression in tendinopathic tendons [165], while differential expression of type I collagen in anterior cruciate ligaments is associated with the degree of healing after a rupture [117]. These findings highlight the potential differences in the genetic and biochemical components contributing to risk of acute and chronic injuries and require further investigation.

In addition to this injury mechanism specific association, it is also interesting to note that the TT genotype is associated with increased risk in a number of other pathologies, such as osteoarthritis, osteoporotic fractures and lumbar disc disease (Table 2.2) [80]. This highlights that certain gene variants may affect the aetiology of phenotypes in different ways. The reasons for these differences remains unknown,

however it may be the result of interactions with other gene variants, interactions with environmental stimuli and/or changes to the mechanical properties of the tissue which may be beneficial or harmful depending on the context.

Table 2.4. A summary of the associations investigated between *COL1A1* rs1800012 and musculoskeletal soft tissue injuries.

Musculoskeletal Soft Tissue Injury	Associated	Studies	Ref
ACL Ruptures	Yes	3	[53;152]
Shoulder Dislocations	?	1	[91]
Achilles Ruptures	?	1	[155]
Achilles Tendinopathy	No	1	[155]
Lateral Epicondylitis (Tennis Elbow)	No	1	[51]

? These studies show tendencies towards an association but larger cohorts are needed to determine whether a true association exists.

2.3.2 Type V Collagen

As mentioned previously, type V collagen is known to interact with types I and III collagen and plays a role in the regulation of collagen fibrillogenesis [18;201]. The rs12722 T/C variant within the 3'-untranslated region (UTR) of the *COL5A1* gene, which encodes the α1 chain of type V collagen, was the first collagen gene variant reported to be associated with Achilles tendon injuries [131]. Specifically, the CC genotype was significantly over-represented in physically active healthy control participants with no self-reported history of Achilles tendon injury when compared to participants with clinically diagnosed chronic Achilles tendinopathy in independent South African and Australian cohorts (Figure 2.7) [131;179]. Several other variants, rs71746744 (-/AGGG), rs169399 (ATCT/-) and rs1134170 (A/T), within the *COL5A1* 3'-UTR were also found to independently associate with risk of Achilles tendinopathy in the same Caucasian South African and Australian cohorts [2]. Specifically, the *COL5A1* rs71746744 AGGG/AGGG, rs169399 -/- and rs1134170 TT genotypes were significantly over-represented in the participants with clinically diagnosed Achilles tendinopathy when compared to the controls in both cohorts as well as in the combined cohorts (Figure 2.7) [2].

Interestingly, one study has also investigated a variant, COL5A1_01 G/A, within intron 65 of the equine *COL5A1* gene for an association with superficial digital flexor (SDF) tendinopathy [199]. Specifically, horses with the AA genotype were significantly (p=0.010) over-represented in the SDF tendinopathy group when compared to yard-matched controls [199]. Other equine *COL5A1* variants which were located upstream of the associated polymorphism were however not

independently associated with SDF. Both the human and equine associated variants are located within the 3'-end of the gene suggesting that this region of the *COL5A1* gene within both species could contain important elements directly involved in the aetiology of tendinopathy.

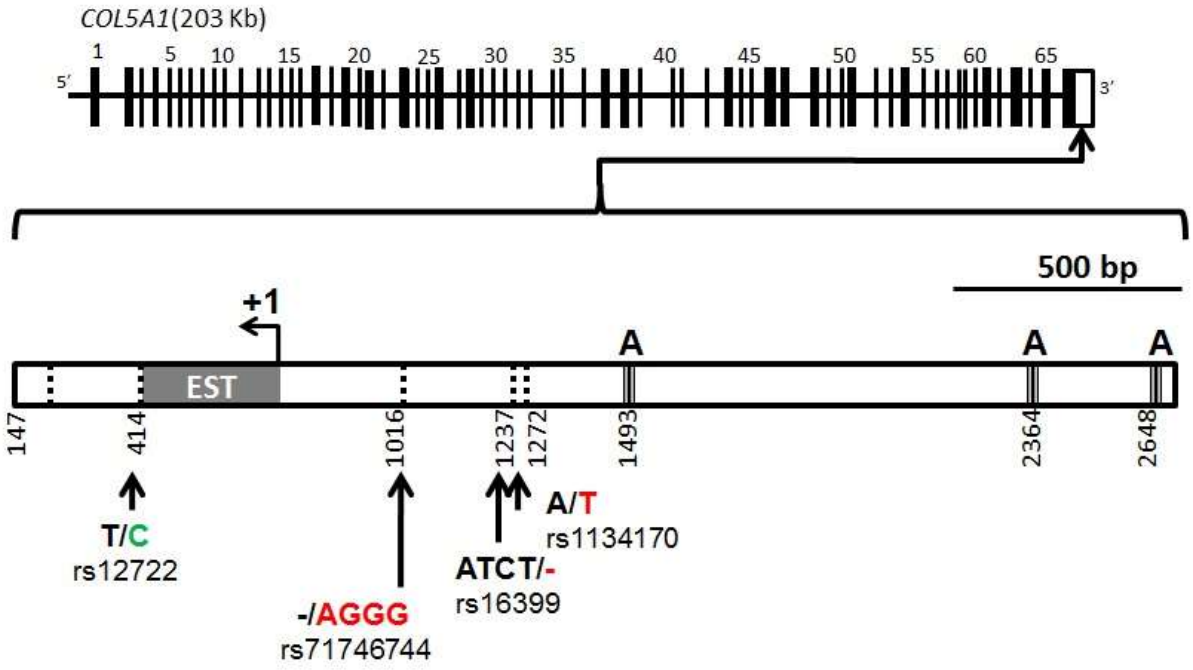


Figure 2.7. A schematic representation of the exon (vertical lines) and intron (horizontal lines) boundaries of the human *COL5A1* gene, which contains 66 exons. Every 5th exon is labelled. Exon 66, which has been enlarged, encodes for the terminal amino acids of the α1(V) chain (black box), the stop codon and the 3'-untranslated region (UTR) (clear box). The four variants (i) rs12722 (T/C), (ii) rs71746744 (-/AGGG), (iii) rs169399 (ATCT/-) and (iv) rs1134170 (A/T) previously shown to be associated with risk of Achilles tendinopathy [2;131;179] and other exercise-related phenotypes [28;29;40;151;153] are all located within the 3'-UTR (indicated by arrows and dotted lines). The alleles in green typeset were associated with reduced risk of Achilles tendinopathy. The alleles in red typeset were associated with increased risk of Achilles tendinopathy. Also noted within the 3'-UTR region are an expressed sequence tag (EST; ENSESTG00000033016, function unknown) and three putative polyadenylation signals (A, multiple lines).

Posthumus et al. [153] further investigated the *COL5A1* rs12722 T/C variant for an association with risk of ACL injuries in a Caucasian South African cohort. The CC genotype was also significantly over-represented in apparently healthy female control participants (27.3%, 23 of 84) when compared with female participants with ACL ruptures confirmed at surgery (5.4%, 2 of 37), but not in a male cohort [153]. The reasons for this gender specific association remain unknown.

Since the 3′-UTR of eukaryotic genes contains elements which are emerging as important post-transcriptional regulators [127;210], it has been suggested that the *COL5A1* 3'-UTR plays a role in the regulation of the *COL5A1* mRNA stability and by implication type V collagen production. In support of this, two major functional allelic forms of the *COL5A1* 3'-UTR, namely the C- and T-forms, have been identified and show significant differences in mRNA stability, with the T-form having increased stability (Figure 2.8) [107]. The T- and C-allelic forms were predominately identified in participants diagnosed with Achilles tendinopathy and asymptomatic controls respectively [107]. The differences in the sequence of the two functional forms of the 3'-UTR is determined by seven variants which include the four variants, rs12722 (T/C), rs71746744 (-/AGGG), rs169399 (ATCT/-) and rs1134170 (A/T), shown to be independently associated with chronic Achilles tendinopathy (Figure 2.8) [107]. Based on these findings, Collins and Posthumus [41] proposed that the relative content of type V collagen in tendons, ligaments and other tissues alters the fibril diameters and packing density within these tissues. This will alter their mechanical properties, and therefore their susceptibility to injury and other exercise-related phenotypes (Figure 2.8). Based on their hypothesis [41], *COL5A1* 3'-UTR variants have also been investigated for associations with other exercise-related phenotypes,

including range of motion (ROM) measurements [1;28;29;40;196] and endurance performance [1;28;151].

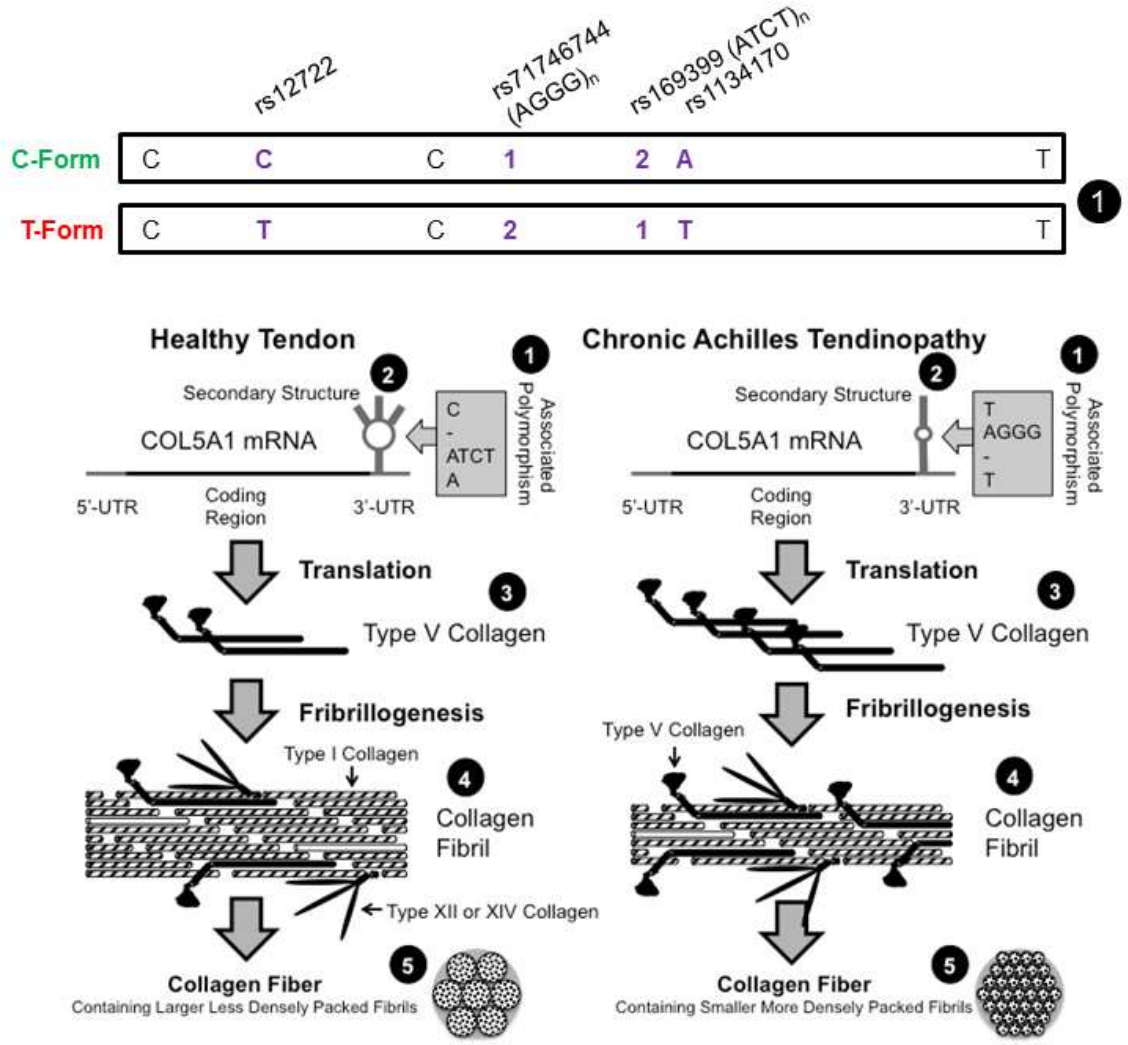


Figure 2.8. A schematic diagram showing the proposed functionality of the *COL5A1* 3'-UTR and its effects on the collagen fibril, adapted from Hay et al. [71]. A functional region with the *COL5A1* 3'-UTR (1), containing a number of gene variants which compose a C- and T-form, is known to alter the secondary structure of the COL5A1 mRNA (2). This altered structure is proposed to result in altered mRNA stability (3 right panel). The T-form is proposed to result in increased levels of type V collagen within the fibril (4 right panel), which results in smaller more densely packed fibrils (5 right panel). Conversely, the C-form is proposed to produce less stable COL5A1 mRNA (3 left panel), resulting in lower levels of type V collagen in the fibril (4 left panel) and therefore larger less densely packed fibrils (5 left panel). The C-form, in green typeset, is associated with reduced risk of Achilles tendinopathy, while the T-form, in red typeset, is associated with increased risk of Achilles tendinopathy.

Mutations within *COL5A1* cause classic EDS, a condition that is characterized by joint hypermobility [124]. Furthermore, it was proposed that *COL5A1* is associated with benign joint hypermobility syndrome [66]. Altered musculotendinous flexibility, which is defined as ‘‘the ability to move a joint through its complete range of motion’’ (ROM), has been cited as an intrinsic risk factor for many injuries [95;96;103], including tendon and ligament injuries. Although flexibility can be improved through regular stretching, it is at least in part also determined genetically [69]. Since variants within *COL5A1* may affect tissue architecture and biomechanical properties [41], variants within this gene were investigated for associations with ROM measurements. The *COL5A1* rs12722 variant was associated with lower limb ROM measurements in a mixed injured/uninjured cohort [40]. In addition, a follow up study in an independent, larger cohort of apparently healthy physically active Caucasian South Africans reported that the *COL5A1* rs12722 CC genotype protected participants against an age-related decline in sit-and-reach ROM [29]. More recently, the *COL5A1* rs12722 CT genotype was also significantly over-represented in a cohort of Italian high-level international rhythmic gymnasts when compared to an active control group, further implicating this variant in modulation of joint mobility [196]. Since this functional variant was associated with sit-and-reach ROM, it was recently investigated for effects on the mechanical properties of human tendons in vivo [58;104]. The results of these studies will be discussed in section 2.5.2.

Reduced ROM, specifically sit-and-reach ROM, has also been associated with improved endurance running economy and performance [45;82]. This relationship is proposed to result from reduced aerobic demand due to increased elastic return and may result in improved performance [82]. Given that the *COL5A1* rs12722 variant is

associated with sit-and-reach ROM [29;40], it is not surprising that this variant was also associated with endurance running performance. The TT genotype of this variant was associated with faster run times in South African Ironman triathletes while triathletes with the CC genotype were over-represented in the slowest tertile for the run stage of this triathlon [151]. The TT genotype was also associated with faster performance in a 56km ultra-marathon road race [28]. Furthermore, when these endurance runners were divided into four quadrants (Fast and inflexible, fast and flexible, slow and inflexible, slow and flexible) based on both their race finishing times and sit-and-reach ROM measures, runners with the *COL5A1* rs12722 T allele were over-represented in the fast and inflexible quadrant.

A second *COL5A1* variant located within the functional *COL5A1* 3'-UTR, rs71746744 (-/AGGG), was later shown to be associated with both endurance running performance and sit-and-reach ROM measures in 56km ultra-marathon runners [1]. Specifically, the rs71746744 AGGG/AGGG genotype was associated with increased performance and reduced sit-and-reach ROM, similar to the rs12722 TT genotype. Interestingly the rs71746744 AGGG and rs12722 T alleles contribute to the functional T-form as described above.

2.3.3 Type XI Collagen

Type XI collagen has been suggested to interact with type V collagen in the regulation of fibrillogenesis during tendon development [207], and sequence variants within the genes encoding the α-chains of type XI collagen have also been investigated for association with Achilles tendinopathy in Caucasian South African

and Australian populations [71]. Although there were no independent genotype associations with risk of chronic Achilles tendinopathy for the *COL11A1* rs3753841 (T/C), *COL11A1* rs1676486 (C/T) and *COL11A2* rs1799907 (T/A) variants, significant gene-gene interactions in modulating risk of Achilles tendinopathy were observed [71]. Specifically, the TCT inferred pseudo-haplotype constructed from rs3753841, rs1676486 and rs1799907 was significantly over-represented in participants with clinically diagnosed Achilles tendinopathy (25.9%) when compared to control participants (17.1%) in a combined Caucasian South African and Australian cohort [71]. Similarly the TCT pseudo-haplotype was significantly over-represented in the Achilles tendinopathy participants when the South African and Australian cohort were analysed separately [71]. Furthermore, the TCT(AGGG) pseudo-haplotype constructed from the type XI collagen genes and independently associated *COL5A1* rs71746744 (-/AGGG) variant was also significantly over-represented in the Achilles tendinopathy participants (25.2%) compared to the controls (9.1%) [71].

The T allele of *COL11A1* rs1676486 results in reduced *COL11A1* expression [159]. The proposed effects of the variants with known function, *COL5A1* rs12722 and *COL11A1* rs1676486, that contribute to the inferred pseudo-haplotypes mentioned above have been previously reported (Figure 2.9) [71]. Based on the functional effects of these variants it was hypothesised that this interaction between *COL11A1* and *COL5A1* may modulate risk of chronic Achilles tendinopathy by affecting mRNA stability, thereby altering type V and/or XI collagen production, which in turn regulates collagen fibril diameter and changes the biomechanical properties of the collagen fibril [71]. In addition, this interaction further corroborates the evidence

supporting integrated roles of type V and type XI collagen in tendon fibrillogenesis [207].

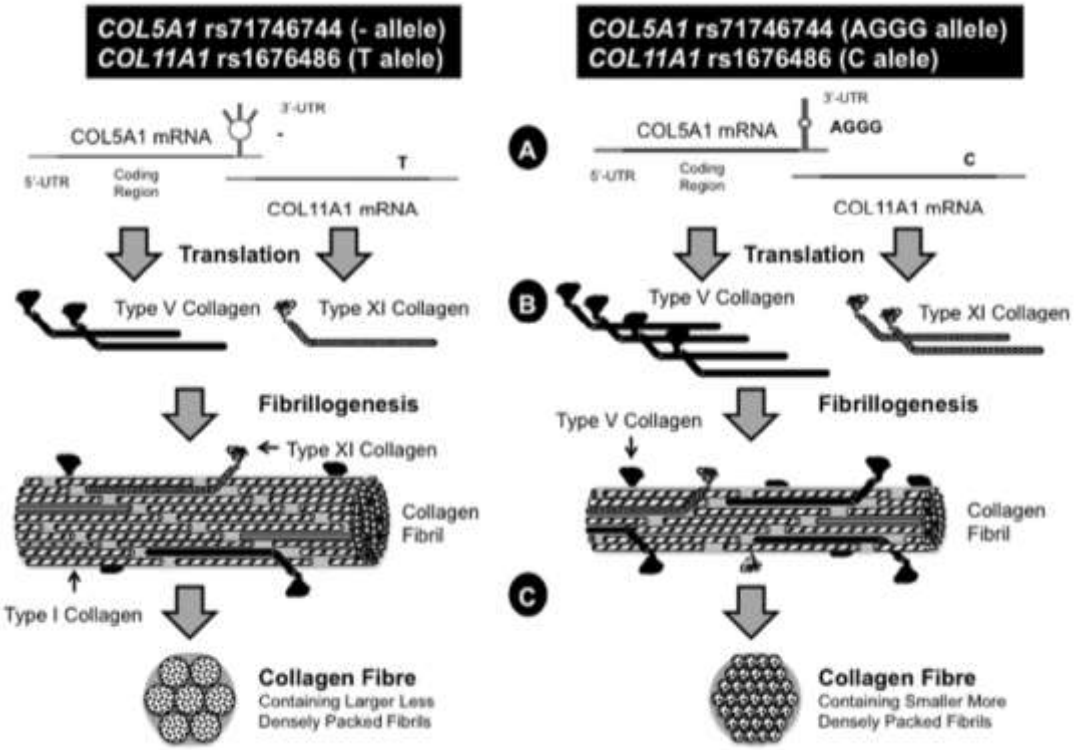


Figure 2.9. A hypothetical schematic diagram illustrating the proposed mechanism of how polymorphisms within *COL5A1* and *COL11A1* potentially affect fibrillogenesis, adapted from Hay et al. [71]. Although there is no evidence that type XI collagen is produced in the mature, diseased and/or healing tendon, it is produced and functional during tendon development.

(A) The *COL5A1* rs71746744 (-/AGGG) and *COL11A1* rs1676486 (C/T) polymorphisms are part of an inferred pseudo-haplotype that is associated with chronic Achilles tendinopathy. The *COL5A1* rs71746744 - allele and the *COL11A1* rs1676486 T allele are both believed to be associated with increased mRNA degradation (increased mRNA degradation is indicated in the left panel, while decreased mRNA degradation is indicated in the right panel).

(B) The altered mRNA stability associated with these polymorphisms is believed to result in altered α1(V) and α1(XI) chain and types V and XI collagen production (decreased in left panel and increased in the right panel).

(C) Types V and XI collagen regulates collagen fibril assembly and diameter and thus the mechanical properties of tendons. There is an inverse relationship between the types V and XI collagen content of the fibril and its diameter. Thinner more densely packed collagen fibrils are produced due to the increased production of types V and XI collagen (right panel). It has previously been proposed that the thinner fibrils are associated with chronic Achilles tendinopathy.

2.3.4 Type XXVII Collagen

As previously mentioned, the fibrillar type XXVII collagen is expressed primarily in cartilage and sites of transition from cartilage to bone, but also in skeletal muscle [23;74;144]. Recently, Saunders et al. [171] investigated variants within the neighbouring *COL27A1* and *TNC* genes as risk factors for chronic Achilles tendinopathy in Caucasian South African and Australian populations. Both of these genes are also situated close to *COL5A1* on chromosome 9 (Figure 2.10), and a variant within the *TNC* gene was previously associated with chronic Achilles tendinopathy [130;171]. Although no independent associations were identified for *COL27A1* rs946053 (G/T), this variant was implicated in a haplotype, with the *TNC* rs13321 and rs2104772 variants, associated with risk of chronic Achilles tendinopathy [171]. This finding highlights the possibility that variants in *COL27A1* and its neighbouring genes also play a role in modulating the risk of chronic Achilles tendinopathy and suggests that *COL27A1* variants may be useful candidates in the investigation of other exercise-related phenotypes.

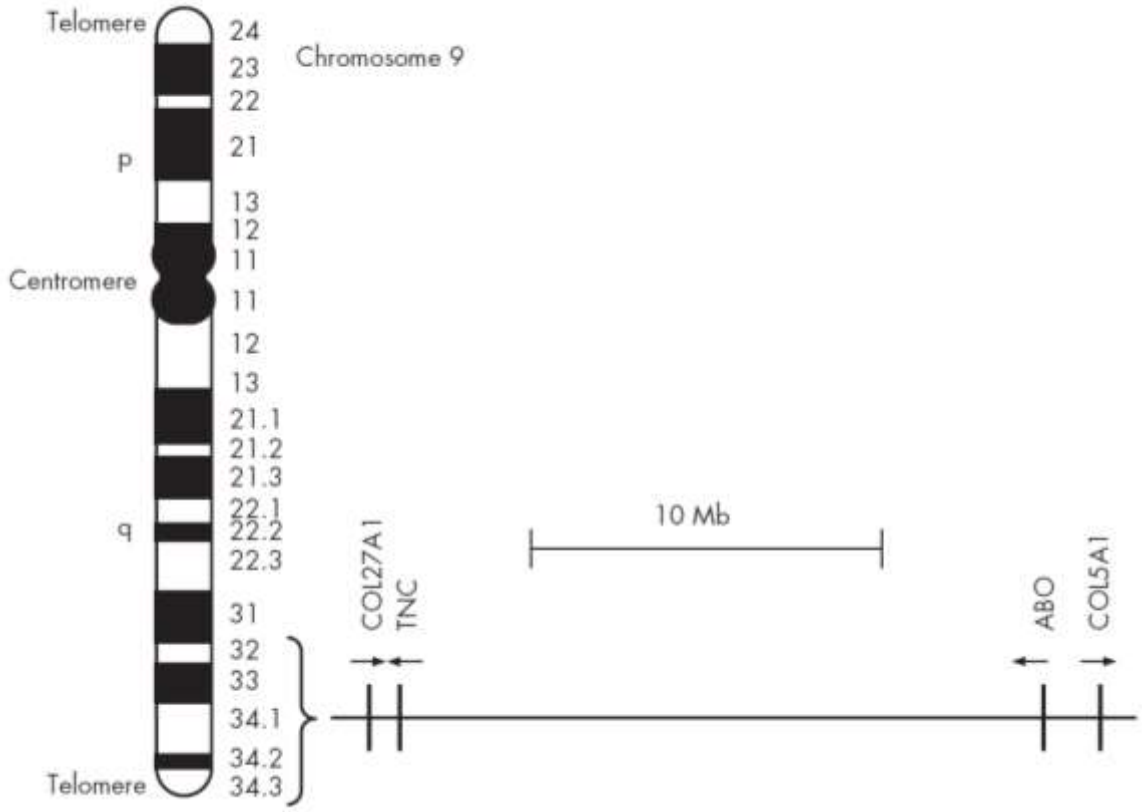


Figure 2.10. A schematic representation of human chromosome 9 showing the relative positions of the *COL27A1*, *TNC*, *ABO* and *COL5A1* genes (adapted from September et al. [182]). The centromere binds with the nuclear spindle during mitosis and meiosis. The telomeres refer to the end regions of the chromosome. The chromosome is comprised of a short arm (p) and a long arm (q). The dark and light areas reflect the unique banding pattern of chromosome 9 when stained using cytogenetic techniques. The nomenclature of each band is indicated on the right side of the chromosome. A segment within the long arm of chromosome 9 (9q32–q34), which encompasses the *COL27A1*, *tenascin C* (*TNC*), *ABO* and *COL5A1* genes, is also shown. The vertical lines represent the relative position of each gene, while the arrows indicate the direction of their transcription. Mb, megabase

The equine *TNC* gene, like the equine *COL5A1* gene as mentioned in section 2.3.2, was also investigated for risk of SDF tendinopathy in horses [199]. The *TNC* BIEC2-696469 T/C variant was found to be associated with SDF tendinopathy. Specifically the CC genotype was significantly (p=0.010) under-represented in SDF tendinopathy

group when compared to yard-matched controls [199]. As in humans, the *COL27A1* gene is upstream of the *TNC* gene on chromosome 25 in horses (http://www.ncbi.nlm.nih.gov/gene/). These results further highlight the potential to investigate the effects of variants on tissues containing these variants in equine models for tendinopathy and other phenotypes.

2.4 GENES ENCODING THE FACITS

2.4.1 Type XII and XIV Collagen

FACIT collagens, such as types XII and XIV collagen, mediate cell-matrix interactions between the collagen fibres and cell surfaces [166], and assist with stabilisation of the attached collagen triple helices by bridging with adjacent fibrils, thereby resisting shear forces [25;164]. Furthermore, these collagens play a similar role to type V collagen in regulating collagen fibril assembly and diameter [213]. The *COL12A1* gene encodes the α1(XII) chains of long (XIIA) and short (XIIB) homotrimeric isoforms of type XII collagen (Figure 2.11). Interestingly the short isoform is predominately expressed in both tendons and ligaments in response to mechanical loading [85]. The α1(XIV) chains of type XIV collagen are encoded for by the *COL14A1* gene (Figure 2.11). As such, September et al. [181] investigated variants within both *COL12A1 and COL14A1* for associations with risk of Achilles tendinopathy in a South African population. However, no significant associations between *COL12A1* rs240736 (T/C), *COL12A1* rs970547 (G/A), *COL14A1* rs4870723 (A/C) and *COL14A1* rs1563392 (T/A) and risk of Achilles tendinopathy were

observed [181]. The *COL12A1* rs970547 and rs240736 variants were also investigated for an association with risk of ACL rupture in a Caucasian South African cohort [154]. Interestingly, a significant association was again only identified after gender stratification of the participants. The *COL12A1* rs970547 AA genotype was significantly over-represented in female participants with clinically diagnosed ACL ruptures when compared to apparently healthy female control participants [154].

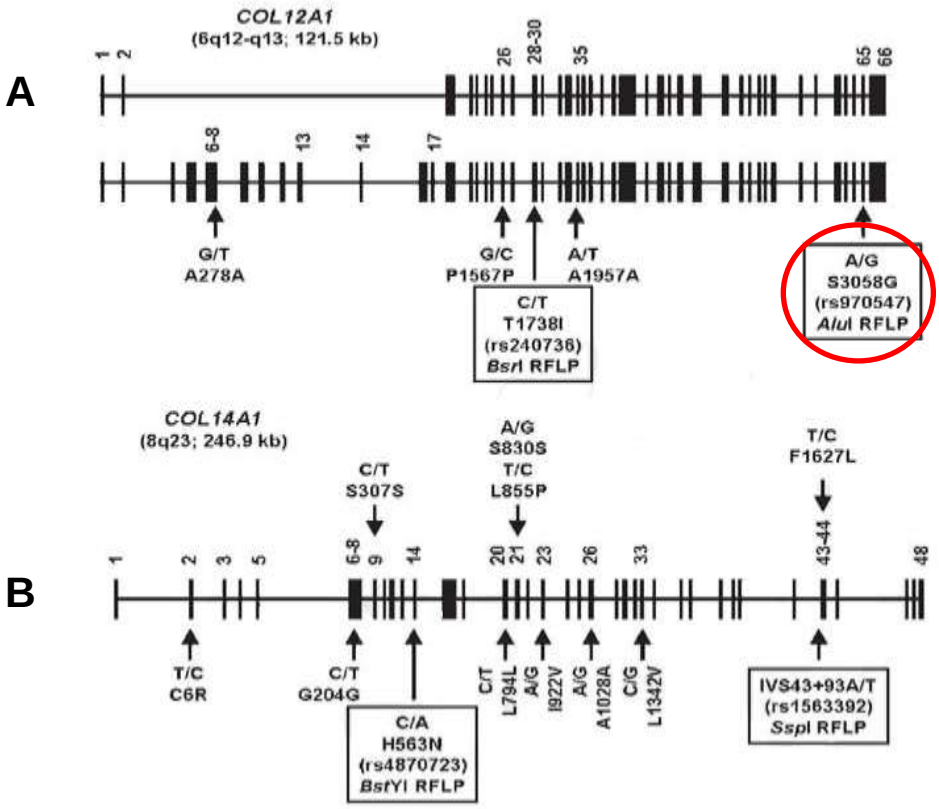


Figure 2.11. A schematic representation of the (A) *COL12A1* and (B) *COL14A1* genes, adapted from September et al. [181]. The *COL12A1* gene has two transcripts, a short (top panel A) and a long (bottom panel A) variant. The exon (vertical lines) and intron (horizontal lines) boundaries are shown. The four variants, *COL12A1* rs240736 and rs970547 and *COL14A1* rs4870723 and rs1563392, investigated by September et al. [181] for associations with Achilles tendinopathy are boxed. The rs970547 associated with risk of ACL ruptures is circled in red [154].

2.5 COLLAGEN FIBRIL ORGANISATION AND THE MECHANICAL PROPERTIES OF MUSCULOSKELETAL SOFT TISSUE

2.5.1 The Collagen Component of Tendons, Ligaments and Connective Tissue Structures in Muscle and other tissues

Collagen fibres (Figure 2.12) form the structural component of tendons, ligaments and connective tissue structures in skeletal muscle. Specifically, these collagen fibres account for approximately 85% of the dry weight of the extracellular matrix (ECM) of tendons, while other components include 1-5% proteoglycans and 2% elastin [83]. Type I collagen fibrils account for 95-99% of the collagen fibres, followed by type III collagen which account for approximately 1-5% (Figure 2.1) [75;83]. Other minor collagens within the tendon include, amongst others, types IV, V, VI, XII and XIV collagen (Table 2.5) [75;83]. Interestingly, expression of these collagens may be specific to portions of the tendon (Table 2.5), including the tendon midsubstance, myotendinous junction (where the tendon connects to the muscle) and osteotendinous junction or entheses (where the tendon connects to the bone). Collagen fbrils come together to form fibres which in turn are bundled together to form fascicles (Figure 2.12) [165]. Each fascicle is surrounded by a loose connective tissue “membrane” called the endotenon [165]. These fascicles are then grouped together, and surrounded by another loose connective tissue “membrane” called the epitenon, to form the tendon (Figure 2.12) [165]. This results in a highly organised structure in which the fibres are all oriented uni-directionally along the long axis of the tendon [75;83].

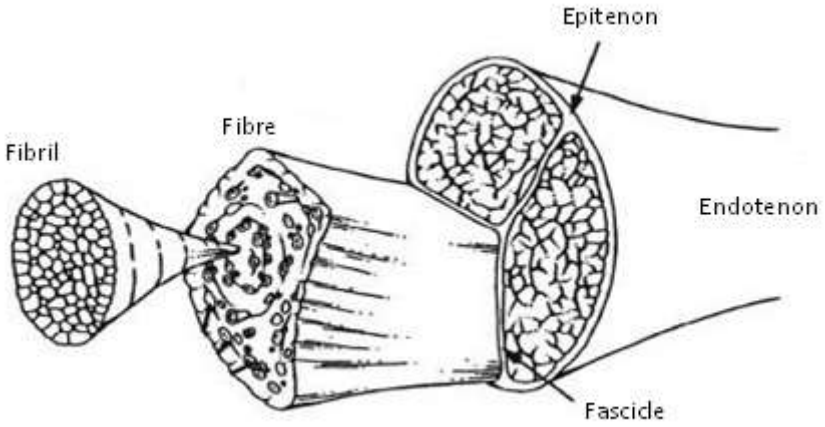


Figure 2.12. The hierarchical structure of a tendon, adapted from Riley [165]. Collagen molecules are grouped together to form the fibril [165]. Fibrils come together to form fibres which in turn are bundled together to form fascicles [165]. Each fascicle is surrounded by connective tissue called endotenon [165]. These fascicles are then grouped together, and surrounded by epitenon, to form the tendon [165]. The hierarchical structure of ligaments is similar to that of tendons as shown above [75].

Although there are important functional differences, ligaments and tendons have a very similar hierarchical structure (Figure 2.12) [75]. Specifically, collagen fibrils form fibres which are then bundled into fascicles, and these fascicles make up the ligament [75]. Despite their similarities in structure the composition of ligaments and tendons is different. The fibres that make up the ligament are less well organised resulting in a multi-directional orientation leading to a weaving pattern [75;83]. Furthermore, differences also exist at a molecular level. Unlike tendons, only 70-80% of the dry mass of ligaments is made up of collagen fibres, while the remainder of the ligament ECM is composed of glycoproteins, proteoglycans and glycosaminoglycans [75]. The collagen composition of ligaments is approximately 90% type I collagen, 10% type III collagen and as much as 12% type V collagen [75]. In addition, the

collagen content of ligaments remains consistent over the entire structure (Table 2.5).

Table 2.5. A summary of "well characterised" collagens identified within tendons, ligaments and the connective tissue structures of skeletal muscle.

Collagen Types	Tendon			Ligament	Skeletal Muscle		
	Midsub	MTJ	OTJ		Endo	Peri	Epi
Fibrillar							
I	√	√	√	√	√	√	√
II			√				
III	√		√	√	√	√	√
V	√			√	√	√	
XI				√			
Non-fibrillar							
IV		√			√		
VI		√	√	√	√		
XII	√			√		√	
XIV	√			√		√	

The localisation of collagen types is indicated by "ticks". The tendon is shown as three separate portions; the midsubstance (Midsub), the myotendinous junction (MTJ) and the osteotendinous junction (OTJ). The three separate connective tissue subdivisions of skeletal muscle are shown; the endomysium (Endo), the perimysium (Peri) and the epimysium (Epi).

In contrast to tendons and ligaments, muscle tissue is composed mainly of myofibres (muscle cells) while the intramuscular connective tissue accounts for only 1-10% of the tissue dry mass [79;92]. This intramuscular connective tissue is comprised of three distinct subdivisions, namely the endomysium, perimysium and epimysium

[62]. The endomysium is a load-bearing network that surrounds individual muscle fibres and most likely transmits forces generated by the muscle by shear [62]. The perimysium surrounds bundles of muscle fibres to form fascicles. Although it is not known, it has been suggested that the perimysium may surround a fascicle from the muscles origin to its insertion point due to its structure differing from that of the mesh-like structure of endomysium. Furthermore, the perimysium may also play a role in cellular signalling [62]. Finally the epimysium surrounds the entire muscle and is continuous with the tendon at the myotendinous junction [62]. Unlike the endo- or perimysium, this subdivision of muscle connective tissue is also known to have larger collagen bundles with a similar orientation and organisation to that seen in tendons [62]. Interestingly, the collagen content within these three subdivisions has also shown to be different (Table 2.5).

In addition to tendons, ligaments and the connective tissue within skeletal muscle, collagens are also present, in varying quantity, within the ECM which occupies the extracellular space of other tissues. This highlights the importance of the regulation of the expression and/or organisation of the collagen types within these connective tissues.

2.5.2 Collagen Levels and their Effects on Fibril Organisation

As previously mentioned, rare mutations in many collagen genes result in serious musculoskeletal disorders [34;121;122;124;144;147;186;205;218] (Table 2.2), and symptoms of these disorders include, but are not limited to, bone fragility, joint hypermobility, skin hyper-extensibility and abnormal wound healing [124;147]. The nature of these symptoms highlights the importance of collagen in maintaining the normal architecture and mechanical properties of musculoskeletal soft tissues. Although no evidence in humans exists, to better understand these disorders and their symptoms, a number of studies have used murine and in vitro models to determine the effects that collagen gene mutations may have on musculoskeletal soft tissue [22;35;78;133;207;213;217].

Investigation of haploinsufficient *Col5a1 +/-* mice, in comparison to wildtype mice, showed that reduced levels of type V collagen resulted in greater numbers of both larger and smaller collagen fibrils in the dermis [206]. These haploinsufficient mice also showed a reduced dermis fibril density when compared to the wildtype mice [206]. These results were later repeated in murine tendon (Figure 2.13) [207]. In addition, haploinsufficient *Col5a1 +/-* and *Coll11a1 +/-* mice showed a more severe phenotype, with aberrant fibril diameter and density, when compared to wildtype and *Col5a1 +/-* only mice (Figure 2.13) [207]. Furthermore, in vitro assays from day 17 chick embryos also identified reduced collagen fibril diameters when type V collagen levels were increased [19].

Interestingly, *Col6a1 -/-* null mice show similar changes to the collagen fibril architecture as haploinsufficient *Col5a1* +/- mice [78]. Specifically, collagen fibrils with smaller diameters and greater density were observed in this knockout murine model (Figure 2.14) [78]. Furthermore, the average weekly distance run by these *Col6a1 -/-* mice was consistently less than the wildtype mice [21], highlighting that these architectural changes may result in changes in biomechanics.

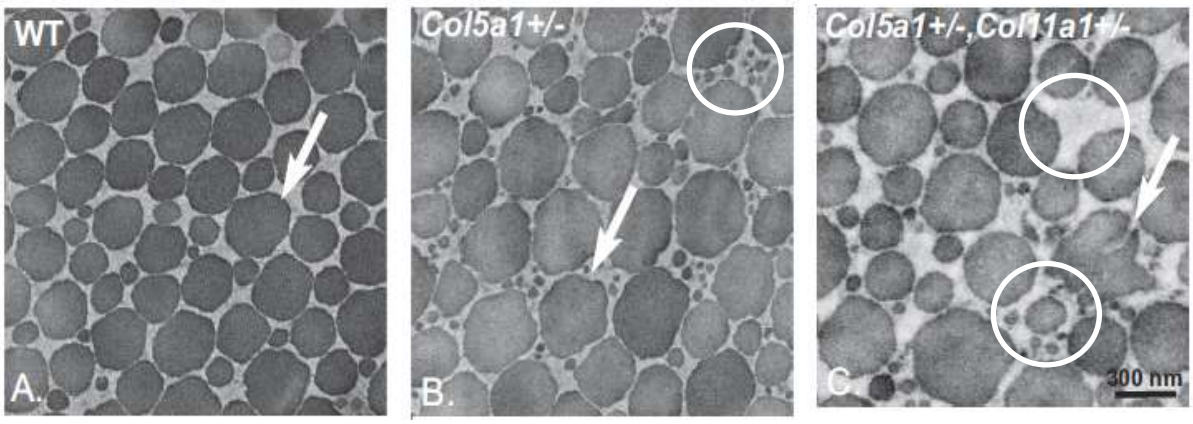


Figure 2.13. Collagen fibril organisation in murine tendon, adapted from Wenstrup et al. [207]. The haploinsufficient *Col5a1* +/- tendons (B) show a greater numbers of larger and smaller collagen fibrils and reduced fibril density (circled in white) when compared to wildtype (WT) tendons (A). Tendons from both haploinsufficient *Col5a1 +/- and Col11a1 +/-* mice (C) showed a more severe phenotype with regard to both fibril diameter and density (circled in white) when compared to both the wildtype (A) and *Col5a1* +/- (B) mice. The white arrows highlight the aberrant structure of the "regular" fibrils in these haploinsufficient murine models when compared to wildtype fibrils.

In vitro assays have also shown that the levels of types XII and III collagen affect the architecture and/or biomechanical properties of the fibril [110;133]. Specifically,

increasing levels of the N-terminal NC-3 domain of type XII collagen reduces inter-fibrillar connections, which may make the matrix surrounding these fibrils more pliable in the absence of cellular stress or more rigid if these domains are able to facilitate cell-mediated alignment and concentration of banded fibrils [133]. These changes to the matrix may then result in fibrils with reduced or larger diameters, respectively [133]. Increased levels of type III collagen has also been shown to reduce fibril diameter in a dose dependent manner [110].

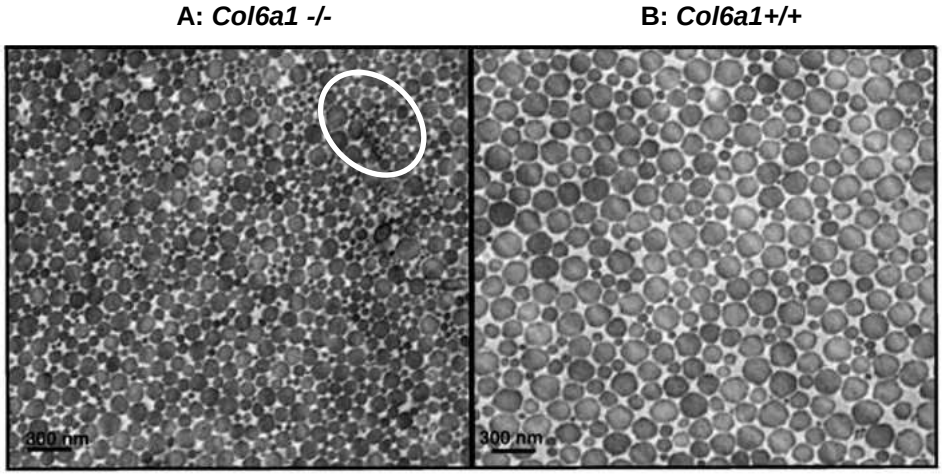


Figure 2.14. Collagen fibril organisation in murine tendon, adapted from Izu et al. [78]. The *Col6a1 -/-* tendons (A) show a greater numbers of smaller collagen fibrils and increased fibril density when compared to the *Col6a1* +/+ tendons (B). The white circle highlights an extreme section of these aberrant small collagen fibrils.

2.5.3 The Effects of Fibril Organisation on the Biomechanical Properties of Musculoskeletal Soft Tissue

Structural and architectural changes to the collagen fibrils are known to alter the biomechanical properties of their containing tissues [102;143]. The measures used to describe the biomechanical properties of these tissues include ultimate tensile strength, Young's modulus and creep [120]. The ultimate tensile strength refers to the maximum stress that the tissue can take before failing [120]. Young's modulus, a measure of stiffness, is defined as the ratio of stress (force per unit area) over strain (deformation relative to initial length) along a particular axis [120]. Finally, creep is the long-term deformation of a tissue due to mechanical stress [120].

An increased mean collagen fibril diameter results in greater ultimate tensile strength [143;145]. Furthermore, the larger diameter also increases the surface area per collagen fibril resulting in an increased number of covalent crosslinks between the adjacent fibrils thereby inhibiting creep [143;145]. A smaller mean collagen fibril diameter results in increased fibril density within the fibre and a reduced surface area per collagen fibril [143;145]. Although the surface area per fibril is reduced, the increased density of smaller fibrils results in a larger overall surface area per unit mass [143;145]. This overall increase in surface area leads to increased creep, and possibly stiffness of the tissue, due to increased electrostatic interactions between the fibrils and ground substance [143].

2.5.4 Effects of the COL5A1 Gene on Fibril Organisation and the Resulting Changes to the Biomechanical Properties of Musculoskeletal Soft Tissue

As mentioned above, murine and in vitro models have shown the effects that collagen gene mutations may have on musculoskeletal soft tissue, including aberrant

changes to collagen fibril size or diameter and fibril density. Therefore, it may be proposed that less severe changes to the musculoskeletal soft tissue architecture may result in other phenotypic effects. For example, changes to the fibril architecture and organisation as a result of functional variants within the *COL5A1* gene may explain associations identified between this gene and musculoskeletal soft tissue injuries and other exercise-related phenotypes (Figure 2.15) [41]. Specifically, smaller fibrils with greater density may result in reduced ROM, improved endurance running performance and reduced risk of musculoskeletal soft tissue injuries [41]. In support of this, one recent in vivo analysis of *COL5A1* rs12722 (T/C) in Japanese males showed that individuals with the CC genotype had greater mean maximal tendon elongation (TT+TC 21.1 ± 5.4 mm, CC 24.5 ± 5.4 mm) and strain (TT+TC 6.51 ± 1.58 %, CC 7.61 ± 1.62 %), and lower stiffness (TT+TC 78.2 ± 18.5 $N.mm^{-1}$, CC 66.2 ± 19.3 $N.mm^{-1}$), for their knee extensors when compared to individuals with the TT or TC genotypes [104]. No significant differences were recorded for the plantar flexors [104]. This study highlights that the level of effect of gene variants may differ between tissue structures, possibly due to differences in environmental stimuli, and adds further evidence to the proposed genetic continuum model updated in figure 2.15 [41;104].

A second in vivo investigation in male and female Caucasian participants identified no associations between the *COL5A1* rs12722 genotypes and patellar tendon or plantar flexor architecture and mechanical properties, including maximal elongation, strain and stiffness [58]. Interestingly, although not significant, the median maximal stiffness showed a tendency to be higher in the TT genotype group (TT 855.7 $N.mm^{-1}$), followed by the TC genotype (TC 707.6 $N.mm^{-1}$), while the CC genotype group

(CC 555.3 $N.mm^{-1}$) reported the lowest maximal stiffness. Differences between the combined TT+TC genotypes compared to the CC genotype group were however not reported. A number of methodological disparities between these studies may explain the differences observed. These include ethnic and gender differences between study populations, as well as the techniques and methods used to determine the dimensional and functional properties of the tendons in vivo [58;104]. Further studies are therefore required to determine the exact effects that *COL5A1* rs12722 may have on tendon architecture and mechanical properties.

Despite the differences in these in vivo tendon analyses, it is known that altered type V collagen expression affects collagen fibril architecture and organisation [206;207]. Like type V collagen, changes to types XII, III and VI collagen levels also result in changes to the fibril architecture and organisation which may affect the biomechanical properties of tissues containing these proteins [21;78;110;133]. Therefore, it may be proposed that common functional variants, like *COL5A1* rs12722, within the genes that encode these proteins may be associated with musculoskeletal soft tissue injuries and other exercise-related phenotypes.

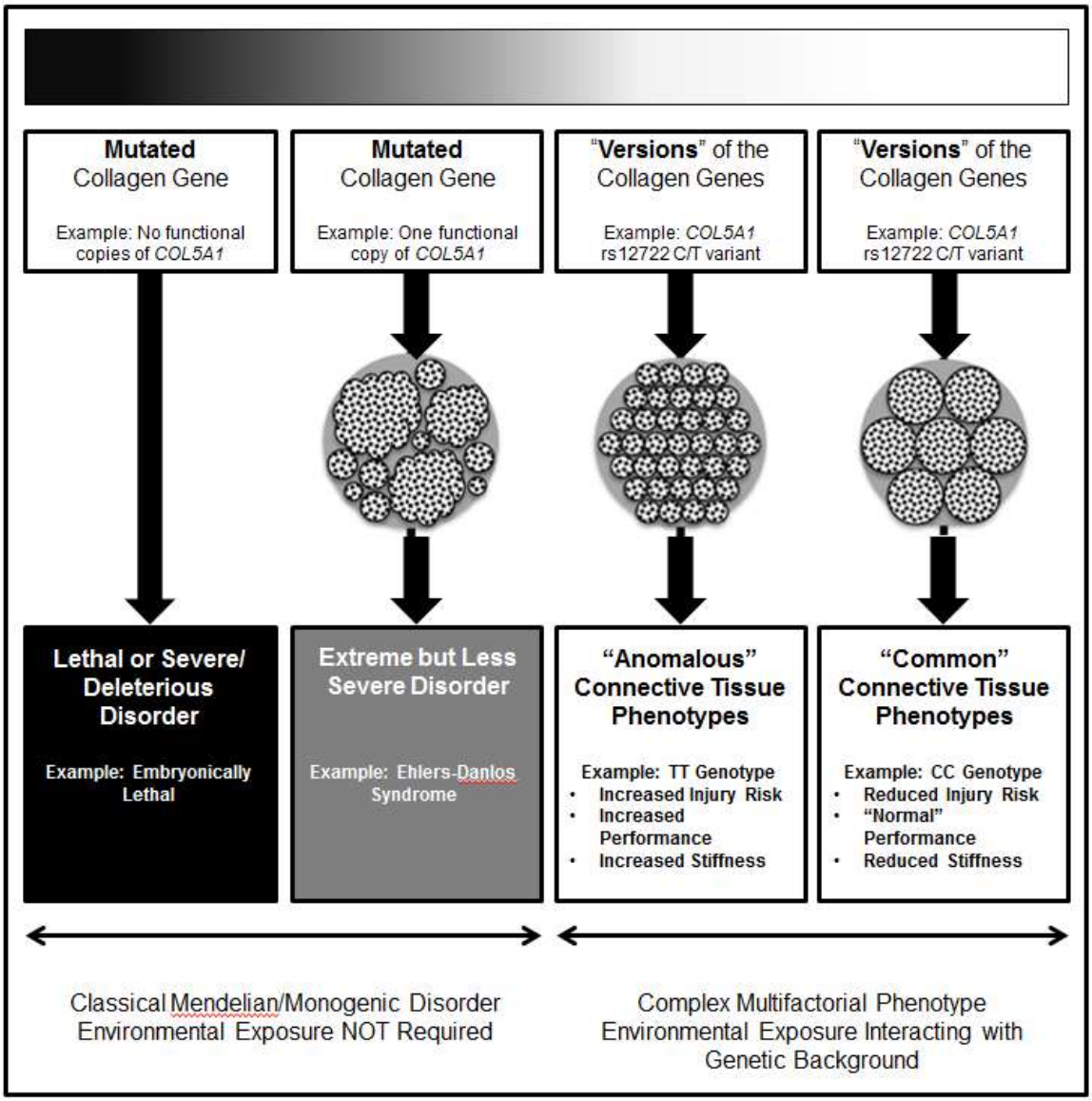


Figure 2.15. A proposed general genetic continuum for collagen genes, adapted from Collins and Posthumus [41]. The black shading represents the lethal or severe phenotypes due to mutations in the collagen genes. At this end of the continuum a single mutation results in the disorder. Murine and in vitro models have shown the effects that collagen gene mutations may have on musculoskeletal soft tissue, including aberrant changes to collagen fibril size or diameter and fibril density [207]. The white shading represents the most beneficial "versions" of the collagen genes. At this end of the continuum variants within collagen genes collectively contribute to the aetiology of the phenotype. Smaller fibrils with greater density, as result of variants such as rs12722, may result in reduced ROM, improved endurance running performance and increased risk of musculoskeletal soft tissue injuries [41;104].

2.5.4 Additional Exercise-Related Phenotypes

As mentioned in section 2.3.2, the *COL5A1* rs12722 gene variant was initially associated with risk of both Achilles tendinopathy [131;179] and ACL rupture [153]. Furthermore mutations within the *COL5A1* gene are known to cause classic type EDS [124]. Due to the characteristics of EDS, joint hypermobility and hyper-extensibility, additional studies investigated and identified associations between *COL5A1* gene variants and ROM [1;28;29;40], and later endurance performance [1;151] (Section 2.3.2). It has therefore been proposed that the relative content of type V collagen in tendons, ligaments and other tissues alters the fibril diameters and packing density within these tissues, and may alter their mechanical properties, and therefore their susceptibility to injury and other exercise-related phenotypes [41]. Investigation of additional candidate exercise-related phenotypes is required to further explore these proposed biomechanical effects.

The altered neuromuscular control hypothesis implicates an increased exercise intensity or duration [174;175;183], and possible tissue damage [174] as primary factors for the aetiology of exercise-associated muscle cramps (EAMC). As mentioned in section 2.3.2, the *COL5A1* rs12722 gene variant was associated with increased endurance running performance [1;151] and musculoskeletal soft tissue injuries [131;153;179], including tendon and ligament injuries. Recently, pre-race serum creatine kinase activity (a marker of muscle damage) tended to be higher in the 20 athletes who developed EAMC during participation within the 2009 56 km Two Oceans ultra-marathon when compared to 29 non-crampers [175]. Furthermore, cross-sectional studies have reported an association between a positive family

history of EAMC and increased risk of EAMC. A cross-sectional survey study highlighted family history of EAMC as a factor that is associated with a past history of EAMC in 1300 marathon runners [125]. This association was later confirmed in a case-control study of 433 Ironman triathletes where the frequency of a positive family history of EAMC was significantly more common in triathletes with a past history of EAMC (36.6%) when compared to those with no history of EAMC (16.4%) [183]. However, in two subsequent prospective cohort studies, a positive family history of EAMC was not associated with the risk of developing EAMC during an event in Ironman triathletes [178] and distance runners [175]. Therefore, since an increased exercise intensity or duration, muscle damage and a familial predisposition may play a role in the aetiology of EAMC, this exercise-related phenotype is a good candidate to further investigate *COL5A1* rs12722 and other collagen gene variants.

Similarly, since the *COL5A1* rs12722 gene variant has been associated with range of motion [28;29;40] and endurance running performance [28;151], and may affect levels of the type V collagen α1 chain synthesis and thereby modulate fibrillogenesis and the biomechanical properties of connective tissues [41], it may be hypothesised that *COL5A1* rs12722, and other collagen gene variants, may be a good candidate for associations with other endurance or power based sports in which muscle tendon stiffness may be regarded as an important feature, such as rugby union [41]. A number of studies have reported differences in rugby union player characteristics between playing positions and playing level [157;158;185]. However, these descriptive studies do not help to determine the underlying factors which may predispose individuals to particular playing positions or playing levels. Only two studies have investigated genetic markers for determining individual playing position

[14;15] and playing level [14]. The first of these studies showed that there were differences in measures of leg power between forwards and backs when the players were grouped according to their *ACE* rs5186 (I/D) genotypes [15]. A second study by the same group investigated the functional *ACTN3* rs1815739 (R/X) gene variant as a marker of playing position and level [14]. No associations were identified between *ACTN3* rs1815739 and playing position or playing level [14]. These results highlight the involvement of a genetic component in predisposing individuals to particular playing positions or playing levels. Therefore this exercise-related phenotype may also be a good candidate to further investigate for associations with *COL5A1* rs12722 and other collagen gene variants.

2.6 CANDIDATE COLLAGEN GENES FOR INVESTIGATION

2.6.1 Type V Collagen

Type V collagen is known to interact with types I and III collagen and plays a role in the regulation of collagen fibrillogenesis [18;201]. Mutations within the *COL5A1* gene are known to cause EDS [124] and the functional *COL5A1* rs12722 T/C variant was associated with Achilles tendinopathy [131], ACL ruptures [153], ROM [28;29;40] and endurance running performance [1;151]. Based on these findings a genetic continuum model was proposed by which common functional variants within this gene may contribute to more complex multifactorial phenotypes [41]. The *COL5A1* rs12722 variant, and others within the *COL5A1* gene, is therefore an

excellent candidate to investigate in additional exercise-related phenotypes, such as EAMC and rugby union.

2.6.2 Type XII Collagen

Type XII collagen, a member of the FACIT collagens sub-family, is important for mediating cell-matrix interactions [166], stabilisation of the collagen triple helices [25;164], in regulating collagen fibril assembly and diameter [213], in a similar fashion to type V collagen. The *COL12A1* gene encodes the α1 chain of type XII collagen and mutations within this gene have also been shown to cause EDS [218]. The *COL12A1* rs970547 A/G variant was investigated in the aetiology of Achilles tendinopathy and ACL rupture [154;181], but was only associated with risk of ACL rupture in Caucasian South African females [154]. In addition, bioinformatic functional analysis of *COL12A1* rs970547 shows that the resulting glycine to serine change is potentially damaging to the COL12A1 peptide and may modulate the function of type XII collagen [105]. This variant should therefore be explored in the aetiology of additional exercise-related phenotypes.

2.6.3 Type III Collagen

Type III collagen consists of three α1(III) chains and is a major fibrillar collagen which forms heterotypic fibrils together with type I collagen [9;114]. Similarly to types I, V and XII collagen, type III collagen is also implicated in fibrillogenesis [56;115;141]. The α1 chain of type III collagen is encoded by the *COL3A1* gene. Like *COL5A1* and *COL12A1*, mutations within the *COL3A1* gene also result in EDS

[33;115;124;186;218]. The non-synonymous *COL3A1* rs1800255 A/G variant within this gene is also associated with a mitral valve and pelvic organ prolapse [37;39;94]. Furthermore, the *COL3A1* rs1800255 variant is proposed to be functional [94]. Specifically, the alanine to threonine change at position 698 of the COL3A1 peptide, as a result of *COL3A1* rs1800255, is proposed to affect the tensile strength of type III collagen fibres [94]. Therefore, since type III collagen is implicated in fibrillogenesis like types I, V and XII collagens, it may be proposed that common potentially functional variants within the *COL3A1* gene, such as *COL3A1* rs1800255, may also be associated with musculoskeletal soft tissue injuries and other exercise-related phenotypes.

2.6.4 Type VI Collagen

The function of type VI collagen remains largely unknown, however, it is believed to play a role at the basement membrane [109;201], and is known to interact with type V collagen in the fibril [190]. In additional, rat models have shown co-localisation of type III and VI collagen in lung tissue [5]. Mutations within the gene which encodes the α1 chain of type VI collagen (*COL6A1*) have been shown to cause muscle diseases such as Bethlem myopathy [17;109] and Ullrich congenital muscular dystrophy [17;109;140]. Bonaldo et al. [21] established a model for myopathy in mice by knocking out the gene responsible for production of the α1 (VI) chain, the *Col6a1* gene. The average weekly distance run by these myopathic *Col6a1 -/-* mice was consistently less than the wildtype mice [21]. In addition, these *Col6a1* null mice showed abnormal fibril diameter and density [78], similar to that observed in *col5al* haploinsufficient mice [206] (Section 2.5.2). Sequence variations in the human

COL6A1 gene have to date been associated with a number of multifactorial conditions, such as ossification of the posterior longitudinal ligament (OPLL) [101;191], ossification of the ligamentum flavum (OLF) [101] and diffuse idiopathic skeletal hyperostosis (DISH) [197]. Specifically a variant within intron 32 (rs35796750 T/C) of the *COL6A1* gene was found to be significantly associated with all of these conditions [101;191;197]. The TT genotype of *COL6A1* rs35796750 was associated with increased risk of both OPLL [191] and DISH [197] in independent Japanese populations. This variant occurs near the branch site of intron 32, where the cytosine to thymine transition is proposed to cause aberrant splicing of the COL6A1 mRNA [191]. This potentially functional variant is therfore another good candidate for investigation in the aetiology of musculoskeletal soft tissue injuries and other exercise-related phenotypes.

2.7 AIMS AND OBJECTIVES OF THE THESIS

The *COL5A1* rs12722 and *COL12A1* rs970547 gene variants have been previously associated with risk of ACL ruptures in females [153;154] and/or chronic Achilles tendinopathy [131;181]. The first aim of this thesis was therefore to investigate the *COL3A1* rs1800255 and *COL6A1* rs35796750 gene variants as risk factors for these musculoskeletal soft tissue injuries. The objectives to address this aim were:

To investigate the *COL3A1* rs1800255 and *COL6A1* rs35796750 gene variants as risk factors for ACL ruptures in a South African cohort (Chapter 3).

To investigate the *COL6A1* rs35796750 gene variant as a risk factor for Achilles tendinopathy in independent SA and Australian cohorts (Chapter 4).

Since the *COL5A1* rs12722 variant was previously associated with endurance performance and ROM, the second aim of this thesis was to further investigate the *COL3A1*, *COL6A1*, and *COL12A1* genes for associations with these exercise-related phenotypes. Therefore, the objectives to address this aim were:

To investigate the *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 gene variants as intrinsic factors for ROM in a SA cohort (Chapter 5).
To investigate the *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 gene variants as intrinsic factors for endurance performance in participants of the SA Ironman triathlon (Chapter 6).

Due to the previously mentioned associations between collagen genes and musculoskeletal soft tissue injuries and other exercise-related phenotypes, the third aim of this thesis was the investigation of *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 in novel exercise-related phenotypes, namely exercise associated muscle cramps (EAMC) and rugby union playing level and position. The objectives to address this aim were:

To investigate the *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 gene variants as risk factors for EAMC in a South African (SA) cohort (Chapter 7).
To investigate the *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 gene variants as intrinsic factors for rugby union playing level and position in professional SA Super 15 rugby union players (Chapter 8).

A summary of the first, second and third aims of this thesis are provided in table 2.6.

Since the *COL6A1* rs35796750 has previously been predicted to be functional, the final aim of this thesis was to investigate the expression of the *COL6A1* gene in individuals with known rs35796750 genotypes.

Investigated Phenotype	Candidate Collagen Gene Variant			
	COL3A1 **rs1800255**	***COL5A1*** **rs12722**	***COL6A1*** **rs35796750**	***COL12A1*** **rs970547**
Anterior Cruciate Ligament Rupture	Chapter 3	Prev Pub[a]	Chapter 3	Prev Pub[b]
Achilles Tendinopathy	Chapter 4	Prev Pub[a]	Chapter 4	Chapter 4
Range of Motion	Chapter 5	Prev Pub[a]	Chapter 5	Chapter 5
Endurance Performance	Chapter 6	Prev Pub[a]	Chapter 6	Chapter 6
Exercise Associated Muscle Cramps	Chapter 7	Chapter 7	Chapter 7	Chapter 7
Rugby Union Playing Level and Position	Chapter 8	Chapter 8	Chapter 8	Chapter 8

Table 2.6. A summary of the candidate collagen gene variants and exercise-related phenotypes that will be investigated in this thesis, as well as those which have been previously investigated and published. The chapter in which each phenotype will be investigated for associations with these variants is also shown. Prev Pub, previously published.

[a] Discussed in Section 2.3.2 [1;28;29;40;131;151;153;179]

[b] Discussed in Section 2.4.1 [154]

CHAPTER 3

GENETIC RISK FACTORS FOR ACL RUPTURES

The data presented in this chapter has been accepted for publication in a condensed form in the peer-reviewed article: O'Connell K, Knight H, Ficek K, Leonska-Duniec A, Maciejewska-Karlowska A, Sawczuk M, Stepien-Slodkowska M, O'Cuinneagain D, van der Merwe W, Posthumus M, Cieszczyk P, Collins M. Interactions between Collagen Gene Variants and risk of Anterior Cruciate Ligament Rupture. *European Journal of Sports Science*. InPress.

3.1 INTRODUCTION

Rupture of the anterior cruciate ligament (ACL), which is one of the injuries studied in this thesis, occurs as a result of a non-contact mechanism in approximately 70% of cases [20] and is one of the most severe injuries in sports [27]. The exact aetiology of ACL ruptures remains unknown. However, a number of extrinsic and intrinsic risk factors, including a genetic component, have been identified [67]. Initially familial studies showed that the risk of an ACL rupture was more than doubled in individuals with a first-degree relative (sibling, parent or child) with history of an ACL rupture [57]. More recently, the rare TT genotype of the *COL1A1* rs1800012 (G/T) single nucleotide polymorphism (SNP) was shown to be under-represented in patients with a cruciate ligament rupture in three separate studies [53;91;152]. The *COL5A1* rs12722 (T/C) and *COL12A1* rs970547 (G/A) variants have also been shown to modulate the risk of ACL ruptures in females in a single cohort [153;154]. Specifically, the CC genotype of *COL5A1* rs12722 decreased the risk, while the AA genotype of *COL12A1* rs970547 increased the risk of ACL rupture in females [153;154]. The *COL1A1, COL5A1* and *COL12A1* genes encode for the α1 chains of

types I, V and XII collagen respectively. These collagens make up the majority of the solid component of ligaments (Chapter 2 Figure 2.12) [59] and play important roles in normal collagen fibrillogenesis [141;213].

Other collagens, such as types III and VI, are also present in the solid component of ligaments [59], where they interact with type I, V and XII collagens to facilitate normal collagen fibrillogenesis [115;129]. The α1 chains of types III and VI collagen are encoded by the *COL3A1* and *COL6A1* gene, respectively. The non-synonymous *COL3A1* rs1800255 (G/A) SNP within exon 30 and the intronic *COL6A1* rs35796750 (T/C) SNP are both proposed to be functional [94;191]. The *COL3A1* rs1800255 variant results in an A698T amino acid transition which is proposed to affect collagen fibre assembly and tensile strength [94]. The GG genotype of this variant is associated with increased risk of mitral valve prolapse [39] while the AA genotype is associated with increased risk of pelvic organ prolapse [37;94]. The intronic *COL6A1* rs35796750 variant is located near an intron branch site and is proposed to result in aberrant splicing of the COL6A1 mRNA [191], which may result in altered or aberrant type VI collagen fibrils. In addition, the TT genotype of this this variant has been associated with increased risk of ossification of the posterior longitudinal ligament [101;191] and the ligamentum flavum [101], as well as diffuse idiopathic skeletal hyperostosis [197].

Since it is already shown that *COL1A1*, *COL5A1* and *COL12A1* variants are independently associated with risk of ACL rupture, it may be hypothesised that *COL3A1* and *COL6A1* variants, as well as, collagen gene-gene interactions are also important modulators of ACL rupture risk. The aims of this study within the thesis

were therefore to (i) determine if *COL3A1* rs1800255 and *COL6A1* rs35796750 are independently associated with risk of ACL rupture in a South African cohort, (ii) repeat the gender-specific independent associations of *COL5A1* rs12722 and *COL12A1* rs970547 with risk of ACL rupture in a larger South African cohort and (iii) investigate gene-gene interactions between collagen variants and risk of ACL rupture. Based on previous findings (Chapter 2 Section 2.6), it was hypothesised that the *COL6A1* rs35796750 TT and *COL12A1* rs970547 AA genotypes, in females only, would be associated with increased risk of ACL ruptures. Previous findings also suggest that either of the *COL3A1* rs1800255 genotypes may be associated with increased risk of ACL rupture. In addition, the *COL5A1* rs12722 CC genotype may be associated with decreased risk of ACL injuries in females.

3.2 MATERIALS AND METHODS

This case-control genetic association study was reported using the recommendations outlined in the genetic association study specific STREGA initiative [113], which adds to the STROBE statement checklist for reporting of observational studies in epidemiology [203]. These recommendations were used for all subsequent genetic association studies (Chapters 4 to 8) in this thesis.

3.2.1 Participants

Four hundred and seventy-seven South African self-reported Caucasian were recruited and included in this study. Two hundred and forty-two participants (177 male and 65 female) with surgically diagnosed ACL ruptures (ACL group) were recruited from the Sports Science Orthopaedic Clinic in Cape Town, South Africa, as

previously described [153]. A subset of the ACL group, 129 (91 male and 38 female) participants were previously recruited and have been included in previous genetic association studies (ACL subset 1) [152-154]. A second subset of 113 (86 male and 27 female) additional participants was recruited for this study (ACL subset 2). The exact mechanism of injury, as defined by the American Orthopaedic Society for Sports Medicine classification system [152], was identified in 161 male (91.0%) and 59 female (90.58%) participants. The 56% (n=136) of the participants, 103 male (58.2%) and 33 female (50.8%), who had ruptured their ACL via a self-reported non-contact (NON) mechanism were analyzed in this study as a separate sub-group (NON sub-group). In addition, 235 previously described apparently healthy participants (145 male and 90 female) with no self-reported history of ACL injury were previously recruited as controls (CON group) from sports and recreational clubs within the greater Cape Town area [152-154]. A breakdown of these groups, as well as their subsets, is shown in figure 3.1.

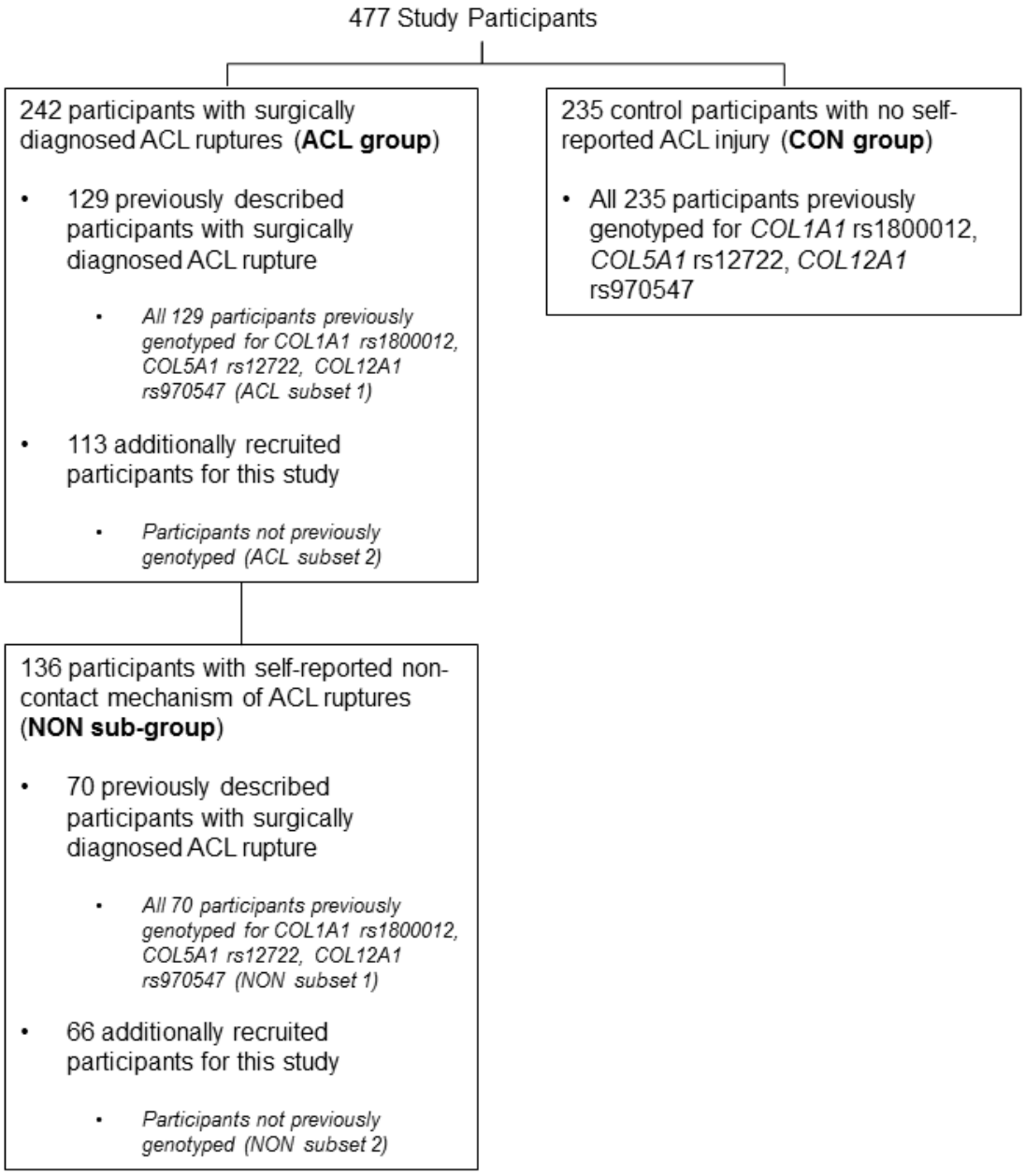


Figure 3.1. A flow diagram showing the breakdown of the South African participants recruited for this study. ACL, Anterior cruciate ligament rupture. CON, control. NON, non-contact mechanism of injury. SA, South African.

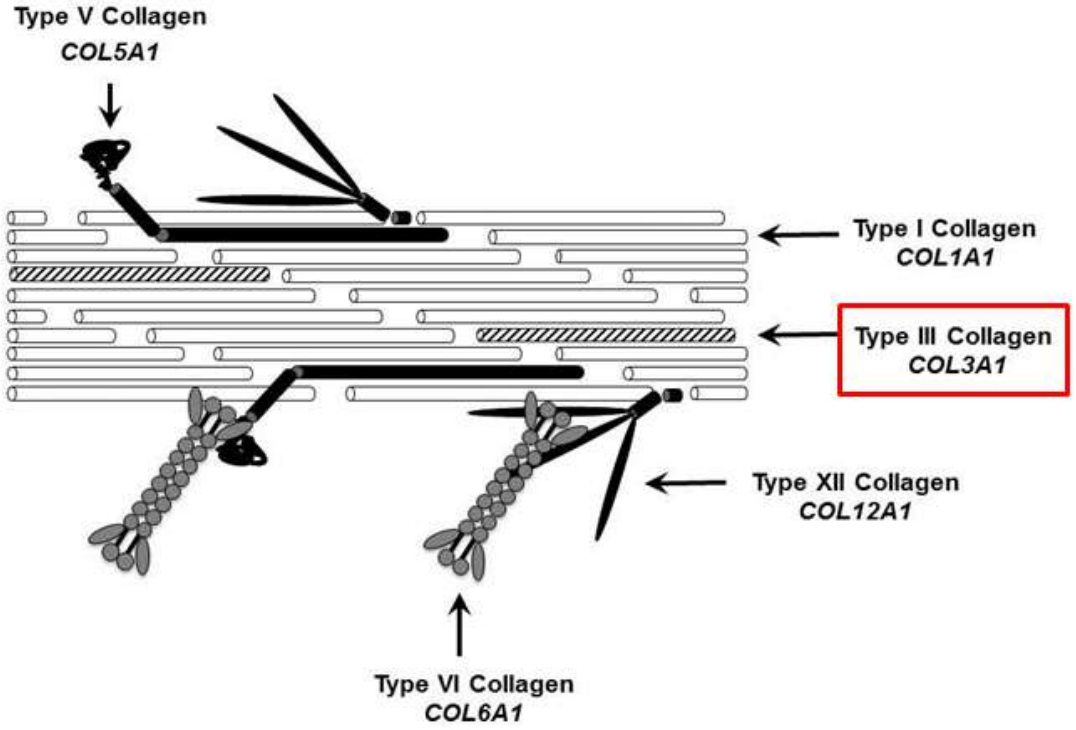
Genotype frequency distributions for *COL5A1* rs12722 and *COL12A1* rs970547 have been published in the 364 previously described South African participants [152-154]. To increase the sample size these variants were genotyped in the additional South African ACL participants recruited for this study.

Prior to participation in this study, all the participants gave informed written consent (Appendix A.3). In addition, each participant completed personal details, medical history, as well as a sports participation questionnaire (Appendix A.4). This study was approved by the Human Research Ethics Committee of the Faculty of Health Sciences within the University of Cape Town, South Africa (Appendix A.1).

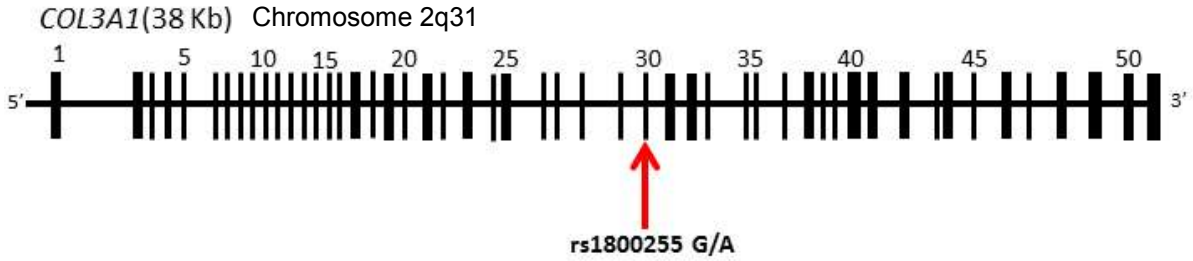
3.2.2 Variant Selection

Four independent sequence variants were investigated in this study. Namely *COL3A1* rs1800255 (Figure 3.2), *COL5A1* rs12722 (Figure 3.3), *COL6A1* rs35796750 (Figure 3.4) and *COL12A1* rs970547 (Figure 3.5). The *COL5A1* rs12722 variant is associated with chronic Achilles tendinopathy [131;179], anterior cruciate ligament ruptures in females [153], range of motion [28;29;40] and endurance running performance [1;151], and the region containing this variant is known to be functional [107]. Variants within the *COL3A1*, *COL6A1* and *COL12A1* genes are associated with a number of soft-tissue phenotypes [39;94;101;154;191;197]. Furthermore, functionality has been proposed for *COL3A1* rs1800255 [94] and *COL6A1* rs35796750 [191], while bioinformatics analysis suggests that *COL12A1* rs970547 may be functional [105].

A



B

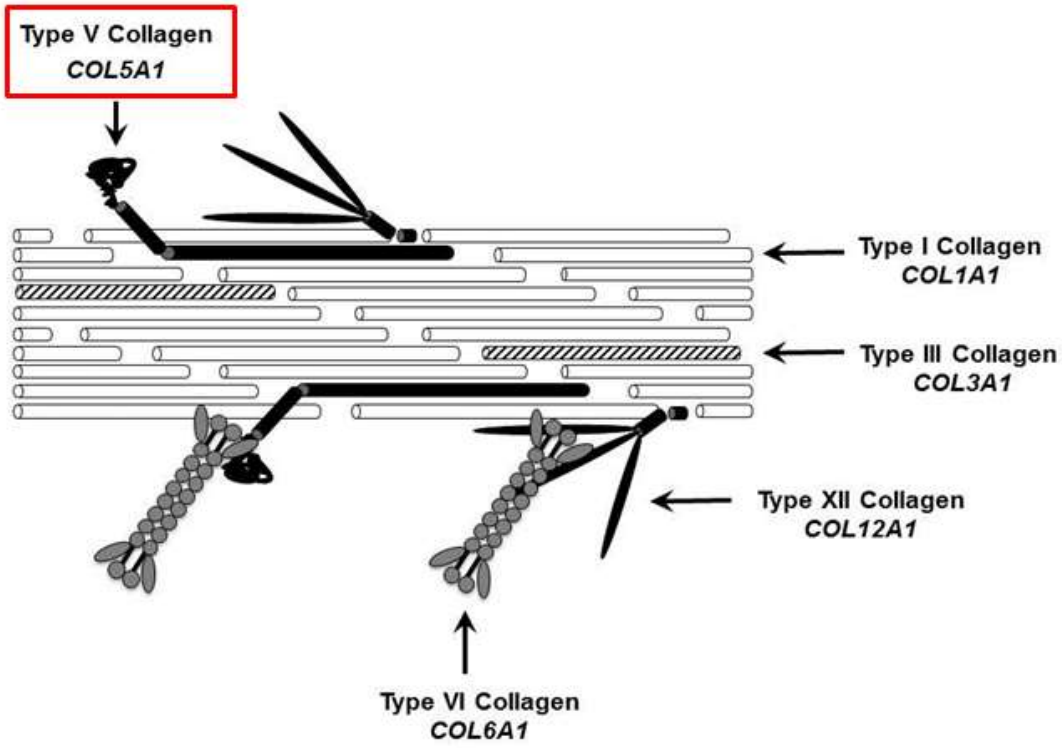


C

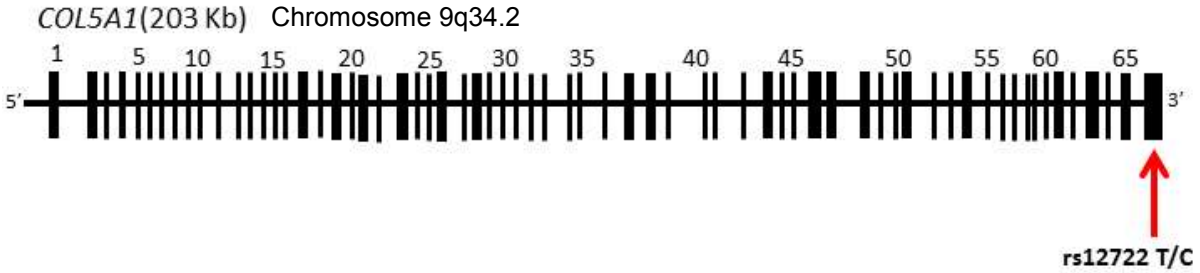
```
                                                    ...TTTTGTCACAGGGTGATGCTGGTGCC 24924
24925 CCTGGTGAACGTGGACCTCCTGGATTGGCAGGGGCCCCAGGACTTAGAGGTGGA 24980
24981 G/A
24982 CTGGTCCCCCTGGTCCCGAAGGAGGAAAGGTAACTCCACAGCATTCCATTCACCT 25037
25038 AGGTTTAAAAAATGCATTTGATTTC...
```

Figure 3.2. Type III Collagen. (A) A schematic diagram showing the position of type III collagen in the fibril. Type III collagen is boxed in red. (B) A schematic diagram of the *COL3A1* gene, as well as the position of the gene variant rs1800255 within the gene (marked by the red arrow). The exon (vertical lines) and intron (horizontal lines) boundaries are shown. Every 5th exon is labelled. (C) The flanking sequence for variant rs1800255 (marked by red text).

A



B

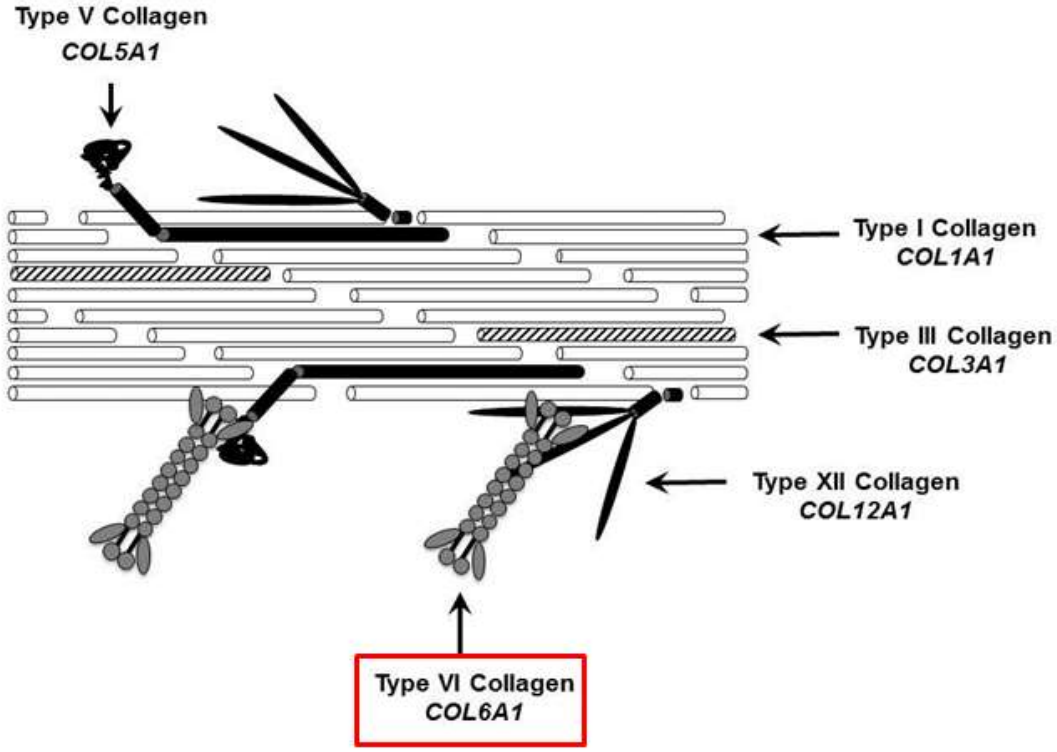


C

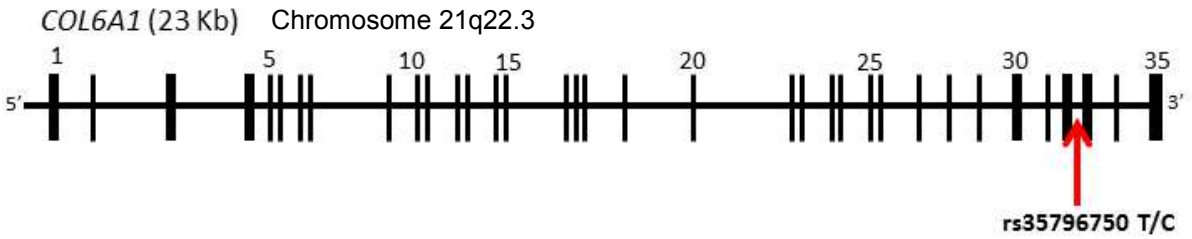
```
                                                  ...CGCGCCCCCACCTGGAGCTGAATCAC 200708
200709 ATGACCTAGCTGCACCCCAGCGCCTGGGCCCGCCCCACGCTCTGTCCACACCCA 200764
200765 T/C
200766 GCGCCCCGGGAGCGGGGCCATGCCTCCAGCCCCCCAGCTCGCCCGACCCATCC 200821
200822 TGTTCGTGAATAGGTCTCAGGGGTTGG...
```

Figure 3.3. Type V Collagen. (A) A schematic diagram showing the position of type V collagen in the fibril. Type V collagen is boxed in red. (B) A schematic diagram of the *COL5A1* gene, as well as the position of the gene variant rs12722 within the gene (marked by the red arrow). The exon (vertical lines) and intron (horizontal lines) boundaries are shown. Every 5th exon is labelled. (C) The flanking sequence for variant rs12722 (marked by red text).

A



B

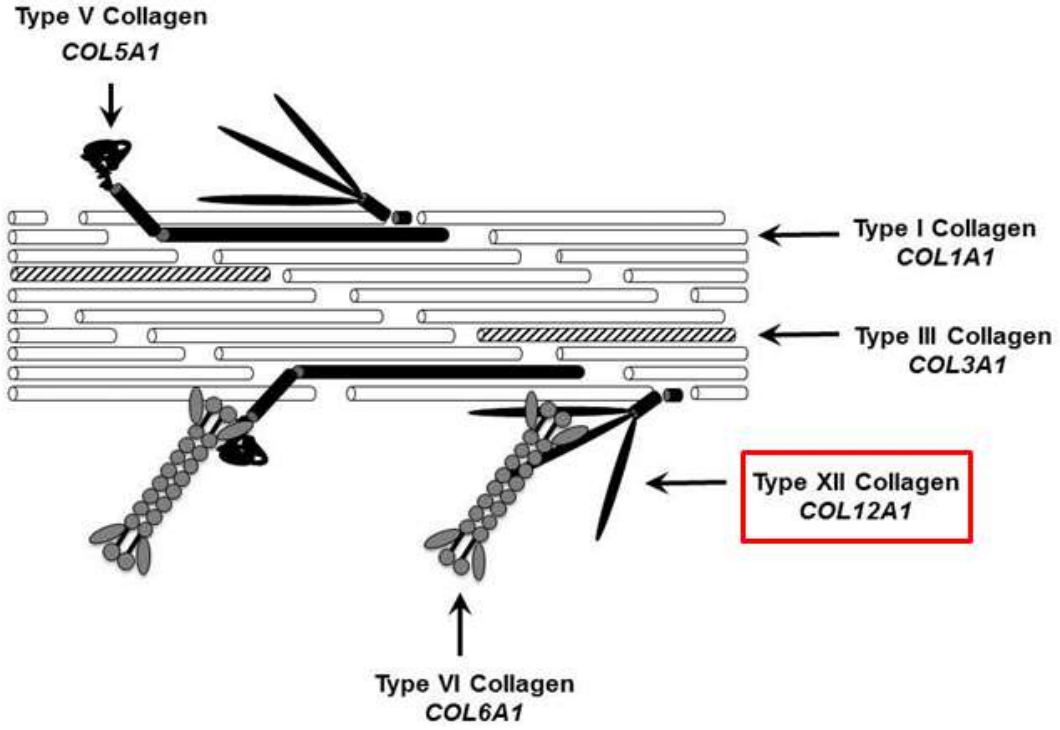


C

```
                                                    ...CAGGCTGCACCTGCCCCGCCTAGGGCG  20 692
20 693  CCCCGCCAGCCAGGGTGGCCTTGTCCCCAGAAAGACGAGGGCAGAGCAGGCTG  20 748
20 749  T/C
20 750  GCCACACCGATACTGTCTGTCCCCACAGGTGGTCTCCGTGGGCATCAAAGACGTG  20 805
20 806  TTTGACTTCATCCCAGGCTCAGACC...
```

Figure 3.4. Type VI Collagen. (A) A schematic diagram showing the position of type VI collagen in the fibril. Type VI collagen is boxed in red. (B) A schematic diagram of the *COL6A1* gene, as well as the position of the gene variant rs35796750 within the gene (marked by the red arrow). The exon (vertical lines) and intron (horizontal lines) boundaries are shown. Every 5th exon is labelled. (C) The flanking sequence for variant rs35796750 (marked by red text).

A



B

C

```
                                                    ...ACTCAGGTATCCGAGGACCCCCAG 118264
118265 GTCCTCCTGGATACTGTGATTCTTCTCAGTGTGCCAGCATCCCATACAACGGGCAA 118320
118321 A/G
118322 GCTATCCAGGTACGTTGTTGCCATTGTAGGAACCCCAACTCACCTGAAGTAAAGAA 118377
118378 CTTTACGGAAATGAAGTTCCTTAG...
```

Figure 3.5. Type XII Collagen. (A) A schematic diagram showing the position of type XII collagen in the fibril. Type XII collagen is boxed in red. (B) A schematic diagram of the *COL12A1* gene, as well as the position of the gene variant rs970547 within the gene (marked by the red arrow). The exon (vertical lines) and intron (horizontal lines) boundaries are shown. Every 5th exon is labelled. (C) The flanking sequence for variant rs970547 (marked by red text).

3.2.3 DNA Extraction and Genotyping Methods

Approximately 4.5 ml of venous blood was collected from all participants by venipuncture of a forearm vein into an ethylenediaminetetraacetic acid (EDTA) vacutainer tube. Blood samples were stored at 4 ˚C until DNA was extracted, as previously described [108], with minor modifications [131]. A brief outline of the methodology is as follows; Blood samples were transferred to 15ml polypropylene tubes and 10ml of TKM1 (10mM Tris-HCl pH 7.6, 10mM KCl, 10mM MgCl2 and 2mM EDTA) buffer containing 2.5% NP40 was added. This lyses the red blood cells. Samples are then incubated and centrifuged at 1200 X g at room temperature for 10 minutes respectively. The centrifugation causes white blood cells (WBC) to form pellets at the bottom the tubes. Following centrifugation the samples were washed with 1 volume of TKM1 buffer. Washed WBC pellets were then resuspended in 800µl TKM2 (10mM Tris-HCl pH 7.6, 10mM KCl, 10mM MgCl2, 0.4M NaCl2 and 2mM EDTA) buffer containing 50µl of 10% SDS and were then incubated for approximately 60 minutes at 55^0C resulting in WBC lysis. The samples were vortexed briefly following the addition of 150µl of 5M NaClO4 and 500µl chloroform. The sample solutions were then transferred to new 1.5ml microfuge tubes and centrifuged at 15000 X g for 5 minutes at room temperature to allow proteins to precipitate. 500µl of the top aqueous layer from each sample was transferred to a new 1.5ml microfuge tube. 1ml of ethanol was then added to the tubes to precipitate the DNA. The precipitated DNA was then pelleted by centrifugation at 15000 X g for 2 minutes at room temperature. The precipitated DNA was then air dried for approximately 30 minutes before being resuspended in 200µl TE buffer (10mM Tris-

HCl, 1mM EDTA, pH 8.0). All tubes were then incubated at 65^0C for 15 minutes in a heating block. DNA was then stored at 4^0C for use.

Genotyping of *COL3A1* rs1800255 and *COL6A1* rs35796750 was performed using custom designed Fluorescence-based Taqman® polymerase chain reaction (PCR) assays (Applied Biosystems, Foster City, CA, USA). Allele specific probes and flanking primer sets (Appendix B) were used along with a pre-made PCR mastermix containing ampliTaq® DNA polymerase Gold (Applied Biosystems, Foster City, CA, USA) in a final reaction volume of 8 µl. The PCR consisted of a 10 min heat activation step (95 ºC) followed by 40 cycles of 15 s at 92 ºC and 1 min at 60 ºC. The PCR reactions were performed on the XP Thermal Cycler, Block model XP-G (BIOER Technology CO., LTD, Tokyo, Japan). Genotypes were determined by end-point fluorescence using a 7900 HT Fast Real-Time PCR System, and SDS Software version 2.3 (Applied Biosystems, Foster City, CA, USA). Genotyping of *COL5A1* rs12722 and *COL12A1* rs970547 was performed using PCR and restriction fragment length polymorphism analysis as previously described (Appendix B) [29;131;153;154]. All genotyping was done at the UCT/MRC Research Unit for Exercise Science and Sports Medicine, University of Cape Town, Cape Town, South Africa.

Investigators were blinded to the phenotypes of the participant samples while genotyping. Two investigators independently confirmed genotyping. Furthermore, a number of positive and negative controls were used to ensure genotyping accuracy. No discrepancies were observed. A total of 433/477 (90.8%), 455/477 (95.4%),

318/364 (87.4%) and 446/477 (93.5%) genotypes were identified for *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547.

3.2.4 Statistical Analysis

Continuous variables were compared between genotype groups using one-way analysis of variance (ANOVA) tests. Chi-squared tests or Fisher tests were used to compare categorical variables. Basic descriptive statistical analysis and frequencies were determined using STATISTICA 11 (StatSoft Inc, Tulsa, Oklahoma, USA). Inferred pseudo-haplotypes between gene variants were tested using Hapstat version 3.0 (Tammy Bailey, Danyu Lin and the University of North Carolina at Chapel Hill, Department of Biostatistics, 3101 McGavran-Greenberg CB #7420, Chapel Hill, North Carolina 27599-7420 USA) [111;112;215]. Hardy-Weinberg equilibrium was determined using the online program Genepop version 4.0.10 (http://genepop.curtin.edu.au/). Significance was accepted when $p<0.05$. No adjustments were made for multiple testing in this study since no obvious appropriate method currently exists [135;146]. The Bonferroni adjustment was considered too conservative since the statistical tests are all on the same group of participants [135]. Adjustment for multiple testing was also considered inappropriate in this case since there is an *a priori* hypothesis that the gene variants investigated in this study are associated with the phenotype [146].

3.3 RESULTS

3.3.1 Participant characteristics

There were significantly less males in the CON group when compared to the ACL group and NON sub-group (Table 3.1). Since gender-specific associations have previously been reported for ACL injuries [153;154], the characteristics of the male and female participants were also analyzed separately. The CON group, as well as the male CON sub-group, were significantly younger and weighed less (age and weight at recruitment) than both the ACL and NON groups or sub-groups (age and weight at time of first injury). The female sub-group were however similarly matched for age. Within the ACL group, the male and female participants' age at recruitment were 4.2 ± 8.4 (n=174) and 5.1 ± 9.5 (n=65) years after the age of their first self-reported ACL rupture, respectively. When appropriately co-varied, the CON group, as well as both the gender sub-groups, were similarly matched for height and BMI when compared to the ACL and NON groups or sub-groups. When co-varied for gender and age at recruitment the CON group weighed less than the ACL group and NON sub-group. Similarly when co-varied for age at recruitment the male CON sub-groups weighed less than their respective ACL and NON sub-groups. The female sub-groups were however similarly matched for weight. The average weight of the male ACL sub-group was 2.3 ± 13.6 kg (n = 171) greater at recruitment when compared the time of their first ACL rupture. Similarly, the average weight of the female ACL sub-group was 0.8 ± 3.5 kg (n = 65) greater at recruitment when compared the time of their first ACL rupture. With the exception of the female ACL

sub-group, which had significantly less South African-born participants, all the other groups and sub-groups were similarly match for country of birth.

The CON group, as well as the male and female CON sub-groups, participated in contact sports and non-contact jumping sports for significantly less years than their respective ACL and NON groups and sub-groups (Table 3.2). There were no significant differences in years of participation in non-contact non-jumping sports between any of the CON groups or sub-groups and their respective ACL and NON groups or sub-groups (Table 3.2). The CON group, and male CON sub-group, participated in skiing for significantly less years than their respective ACL and NON groups and sub-groups (Table 3.2). There were no differences between the female CON and ACL or NON sub-groups and years of participation in skiing (Table 3.2). A list of reported contact and non-contact sports can be found in appendix C.3.

There were no genotype effects on any of the participant characteristics for *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 (Table 3.3).

Table 3.1. General characteristics for the South African participants with anterior cruciate ligament rupture (ACL group), non-contact mechanism of injury (NON sub-group) and apparently healthy controls (CON group) recruited for this study. The male and female sub-groups are also shown.

All Participants	**CON (235)**	**ACL (242)**	**p-value[a]**	**NON (136)**	**p-value[b]**
Gender (% male)	61.7 (145)	73.1 (177)	**0.008**	75.7 (103)	0.**008**
Age (yrs)	29.3 ± 11.2 (229)	26.8 ± 11.0 (211)	**0.018**	26.2 ± 10.3 (129)	0.292
Height (cm)	175.0 ± 9.4 (229)	177.4 ± 9.7 (221)	0.332[c]	178.2 ± 9.1 (129)	0.292[c]
Weight (kg)	74.1 ± 14.8 (230)	80.3 ± 17.0 (221)	**0.004[d]**	80.4 ± 16.2 (133)	**0.018[d]**
BMI	23.9 ± 4.1 (229)	24.7 ± 5.8 (221)	0.242[d]	24.7 ± 5.4 (129)	0.431[d]
South African Born (%)	82.3 (195)	82.1 (183)	0.118	82.7 (110)	0.118
Male Participants	**CON (145)**	**ACL (177)**	**p-value[a]**	**NON (103)**	**p-value[b]**
Age (yrs)	29.5 ± 11.8 (142)	25.8 ± 10.0 (151)	**0.003**	25.3 ± 8.9 (96)	**0.003**
Height (cm)	180.5 ± 6.4 (142)	181.3 ± 6.9 (163)	0.273	181.7 ± 6.7 (101)	0.165
Weight (kg)	81.6 ± 13.0 (141)	86.5 ± 14.7 (163)	**0.002[e]**	86.2 ± 13.5 (100)	**0.007[e]**
BMI	24.6 ± 4.6 (141)	25.6 ± 5.5 (163)	**0.023[e]**	25.4 ± 5.8 (101)	0.148[e]
South African Born (%)	87.8 (122)	85.3 (139)	0.871	82.2 (83)	0.575
Female Participants	**CON (90)**	**ACL (65)**	**p-value[a]**	**NON (33)**	**p-value[b]**
Age (yrs)	28.9 ± 10.2 (87)	29.4 ± 12.8 (60)	0.812	29.1 ± 13.3 (33)	0.943
Height (cm)	166.1 ± 5.9 (87)	166.3 ± 7.7 (58)	0.919	166.8 ± 6.1 (31)	0.877
Weight (kg)	62.4 ± 8.5 (89)	63.0 ± 9.6 (58)	0.739[e]	62.6 ± 9.4 (33)	0.904[e]
BMI	22.6 ± 2.7 (87)	22.1 ± 5.7 (58)	0.449[e]	22.5 ± 3.0 (31)	0.849[e]
South African Born (%)	83.9 (73)	73.3 (44)	**0.014**	84.4 (27)	0.551

Values are expressed as mean ± standard deviations or as percentages where appropriate. Number of participants is indicated in parentheses. Values in bold typeset are significant ($p<0.05$). For the ACL and NON groups and sub-groups age, height, weight and BMI are reported from time of injury. For the CON group and sub-groups the age, height and BMI are reported from time of recruitment. NON, non-contact mechanism of injury.

[a] CON vs. ACL
[b] CON vs. NON
[c] Co-varied for gender
[d] Co-varied for gender and age at recruitment
[e] Co-varied for age

Table 3.2. Years of participation in contact, non-contact non-jumping, non-contact jumping sports, as well as skiing, for the South African participants with anterior cruciate ligament rupture (ACL group), non-contact mechanism of injury (NON sub-group) and apparently healthy controls (CON group) recruited for this study. The male and female sub-groups are also shown.

All Participants	**CON (235)**	**ACL (242)**	**p-value[a]**	**NON (136)**	**p-value[b]**
Contact Sports (yrs)	6.9 ± 8.2 (166)	12.3 ± 7.6 (152)	**<0.001**	12.0 ± 7.3 (91)	**<0.001**
Non-contact Non-jumping Sports (yrs)	27.2 ± 21.4 (216)	24.1 ± 20.3 (197)	0.126	25.5 ± 21.7 (117)	0.498
Non-contact Jumping Sports (yrs)	3.5 ± 6.6 (129)	10.0 ± 7.6 (62)	**<0.001**	11.7 ± 8.4 (32)	**<0.001**
Skiing (yrs)	1.8 ± 7.5 (115)	8.1 ± 8.4 (17)	**0.002**	13.8 ± 11.1 (4)	**0.002**
Male Participants	**CON (145)**	**ACL (177)**	**p-value[a]**	**NON (103)**	**p-value[b]**
Contact Sports (yrs)	9.2 ± 8.6 (115)	12.5 ± 7.6 (145)	**0.001**	12.1 ± 7.2 (88)	**0.012**
Non-contact Non-jumping Sports (yrs)	27.0 ± 21.4 (134)	23.2 ± 20.1 (146)	0.127	24.6 ± 21.2 (89)	0.411
Non-contact Jumping Sports (yrs)	1.2 ± 3.7 (68)	8.8 ± 9.2 (31)	**<0.001**	11.6 ± 10.7 (16)	**<0.001**
Skiing (yrs)	1.7 ± 6.3 (70)	10.0 ± 9.7 (11)	**<0.001**	16.7 ± 11.5 (3)	**<0.001**
Female Participants	**CON (90)**	**ACL (65)**	**p-value[a]**	**NON (33)**	**p-value[b]**
Contact Sports (yrs)	1.6 ± 3.1 (51)	7.4 ± 7.3 (7)	**<0.001**	8.0 ± 8.9 (3)	**0.003**
Non-contact Non-jumping Sports (yrs)	27.6 ± 21.6 (82)	26.5 ± 20.9 (51)	0.779	28.5 ± 23.2 (28)	0.840
Non-contact Jumping Sports (yrs)	6.0 ± 8.0 (61)	11.2 ± 5.4 (31)	**0.002**	11.8 ± 5.6 (16)	**0.009**
Skiing (yrs)	2.1 ± 9.2 (45)	4.5 ± 3.3 (6)	0.526	5.0 ± 0.0 (1)	0.754

Values are expressed as mean ± standard deviations. Number of participants is indicated in parentheses. Values in bold typeset are significant ($p<0.05$).

[a] CON vs. ACL

[b] CON vs. NON

Table 3.3. Genotype effects of *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 on physiological characteristics of participants.

Polymorphism	p-values			
	Age Injured	Height	Weight	Gender
***COL3A1* rs1800255**	0.521	0.556	0.602	0.945
***COL5A1* rs12722**	0.799	0.265	0.335	0.457
***COL6A1* rs35796750**	0.585	0.823	0.820	0.993
***COL12A1* rs970547**	0.624	0.918	0.950	0.992

3.3.2 Participant and Family History of Soft Tissue Injuries

Family history, which includes any blood relative, of any ligament injury reported by the participant at the time of recruitment was significantly higher in the ACL group (odds ratio (OR) = 1.8, 95% confidence interval (CI) = 1.2 – 2.6, p=0.004) and NON sub-group (OR = 1.9, 95% CI = 1.2 – 3.0, p=0.004) when compared to the CON group (Table 3.4). A family history of ACL injuries was also similarly significantly higher in the ACL group (OR = 9.1, 95% CI = 3.5 – 23.5, p<0.001) and NON sub-group (OR = 11.4, 95% CI = 4.3 – 30.4, p<0.001) when compared to the CON group (Table 3.4). There were no significant differences in the family history of any tendon injuries between the groups.

Individually, significantly more participants in the ACL group and NON sub-group reported previous injuries of any ligament, as well as, previous posterior cruciate ligament (PCL), lateral collateral ligament (LCL) and/or the medial collateral ligament

(MCL) injuries in the knee (Table 3.4). There were significantly more self-reported participants with a history of lateral and/or medial ankle ligament injuries within the NON sub-group, but not within the ACL group. Finally the groups and sub-groups were similarly matched for a history of joint capsule disease and any tendon injury. Similar results were observed when male and female participants were analysed separately (Appendix C.1).

Table 3.4. Participant and family history of soft tissue injuries for the participants recruited for this study.

	CON (235)	ACL (242)	p-value[a]	NON (136)	p-value[b]
Participant previous ligament injury	34.0 (80)	43.8 (106)	**0.029**	43.4 (59)	**0.014**
Knee ligament injury[c]	3.0 (7)	11.6 (28)	**0.003**	11.0 (15)	**0.007**
Ankle ligament strain[d]	21.7 (51)	28.9 (70)	0.070	31.6 (43)	**0.039**
Participant previous tendon injury	20.0 (47)	21.5 (52)	0.689	25.7 (35)	0.640
Participant history of joint capsule disease	9.8 (23)	10.3 (25)	0.322	13.2 (18)	0.111
Family history ligament injury	26.8 (63)	39.3 (95)	**0.004**	41.2 (56)	**0.003**
ACL injury	2.1 (5)	16.5 (40)	**<0.001**	19.9 (27)	**<0.001**
Family history tendon injury	8.1 (19)	11.6 (28)	0.202	9.6 (13)	0.122

Values are expressed as percentages with the number of participants (n) indicated in parentheses. Values in bold typeset are significant ($p<0.05$). CON, apparently healthy controls. ACL, anterior cruciate ligament. NON, non-contact mechanism of injury.

[a] CON vs. ACL

[b] CON vs. NON

[c] includes the posterior cruciate ligament, the lateral collateral ligament and the medial collateral ligament.

[d] includes the lateral and medial ankle ligaments.

3.3.3 COL3A1 and COL6A1 Genotype Frequency Distributions

There were no significant differences in the *COL3A1* rs1800255 and *COL6A1* rs35796750 genotype or allele frequency distributions between the ACL group and CON group or NON sub-group and CON group, or their respective gender sub-groups (Tables 3.5 and 3.6). A tendency was however observed for the *COL3A1* rs1800255 minor A allele (p=0.071) and AA genotype (p=0.099) to be under-represented in the male ACL sub-group when compared to the male CON sub-group. Similar genotype frequency distributions were obtained when control participants with a self-reported history of any previous ligament injury were excluded from the analysis (Appendix C.2). Therefore these participants were included to increase the sample size and statistical power of the study. All *COL3A1* (Table 3.5) and *COL6A1* (Table 3.6) groups and sub-groups were all in Hardy-Weinberg Equilibrium (HWE).

Table 3.5. A comparison of the genotype frequency distributions for *COL3A1* rs1800255 (G/A), between control (CON group), anterior cruciate ligament rupture (ACL group) and non-contact mechanism of injury (NON sub-group) groups. Gender specific comparisons between the two groups are also reported.

	***COL3A1* rs1800255 Genotype**			**n**	**Genotype p-value**	**HWE**	**Minor Allele**	**Allele p-value**
	GG	**GA**	**AA**					
All Participants								
CON	52.1 (111)	38.0 (81)	9.9 (21)	213		0.324	28.9 (123)	
ACL	55.9 (123)	39.1 (86)	5.0 (11)	220	0.151[a]	0.476	24.5 (108)	0.150[a]
NON	57.4 (70)	36.9 (45)	5.7 (7)	122	0.368[b]	1.000	24.2 (59)	0.189[b]
Male Participants								
CON	51.6 (65)	38.9 (49)	9.5 (12)	126		0.194	29.0 (73)	
ACL	58.9 (96)	37.4 (61)	3.7 (6)	163	0.099 [a]	0.500	22.3 (73)	0.071[a]
NON	60.2 (56)	34.4 (32)	5.4 (5)	93	0.333 [b]	1.000	22.6 (42)	0.133[b]
Female Participants								
CON	52.9 (46)	36.8 (32)	10.3 (9)	87		0.431	28.7 (50)	
ACL	47.4 (27)	43.9 (25)	8.8 (5)	57	0.694 [a]	1.000	30.7 (35)	0.721[a]
NON	48.3 (14)	44.8 (13)	6.9 (2)	29	nd	nd	29.3 (17)	nd

Values are expressed as percentages with the number of participants indicated in parentheses. The total number (n) of participants genotyped in each individual or combined cohort is also indicated. Values in bold typeset are significant ($p<0.05$). All the ACL ruptures were diagnosed at surgery.

HWE, Hardy-Weinberg Equilibrium. nd, not determined.

[a] CON vs. ACL

[b] CON vs. NON

Table 3.6. A comparison of the genotype frequency distributions for *COL6A1* rs35796750 (T/C), between the control (CON group) participants, previously recruited anterior cruciate ligament rupture (ACL group subset 1) and previously recruited non-contact mechanism of injury (NON sub-group subset 1) groups. Gender specific comparisons between the two groups are also reported.

	***COL6A1* rs35796750 Genotype**			**n**	**Genotype p-value**	**HWE**	**Minor Allele**	**Allele p-value**
	TT	**TC**	**CC**					
All Participants								
CON	33.0 (65)	46.7 (94)	20.3 (40)	199		0.664	44.0 (174)	
ACL subset 1	32.1 (38)	53.6 (64)	14.3 (17)	119	0.345[a]	0.271	41.2 (98)	0.708[a]
NON subset 1	32.8 (22)	55.2 (37)	11.9 (8)	67	0.287[b]	0.215	39.6 (53)	0.399[b]
Male Participants								
CON	33.6 (42)	44.0 (55)	22.4 (28)	125		0.280	44.4 (111)	
ACL subset 1	30.4 (24)	57.0 (45)	12.7 (10)	79	0.118[a]	0.165	41.1 (65)	0.711[a]
NON subset 1	29.2 (14)	58.3 (28)	12.5 (6)	48	0.196[b]	0.241	41.7 (40)	0.646[b]
Female Participants								
CON	31.1 (23)	52.7 (39)	16.2 (12)	74		0.637	42.6 (63)	
ACL subset 1	35.0 (14)	47.5 (19)	17.5 (7)	40	0.867[a]	1.000	41.3 (33)	0.889[a]
NON subset 1	42.1 (8)	47.4 (9)	10.5 (2)	19	nd	nd	28.9 (11)	nd

Values are expressed as percentages with the number of participants indicated in parentheses. The total number (n) of participants genotyped in each individual or combined cohort is also indicated. All the ACL ruptures were diagnosed at surgery. The ACL and NON subset 1 participants were previously recruited by Posthumus et al. [153;154].

HWE, Hardy-Weinberg Equilibrium. nd, not determined due to sample size.

[a] CON vs. ACL

[b] CON vs. NON

3.3.4 COL5A1 and COL12A1 Genotype Frequency Distributions

Since both gender-specific associations have previously reported for the *COL5A1* and *COL12A1* variants [153;154], the male and female sub-groups were analysed separately. When only the males were analysed, there were no significant differences in the *COL5A1* rs12722 genotype frequency distributions between the ACL and CON or NON and CON sub-groups (Table 3.7). Similarly, no significant differences were identified in the *COL12A1* rs970547 genotype distributions between the male ACL and CON or NON and CON sub-groups (Table 3.8). As expected, the *COL5A1* rs12722 CC genotype was significantly (p=0.001 OR = 5.9, 95% CI 1.9 – 17.9) under-represented in the larger female ACL (6.8%, n=4) sub-group when compared to the female CON (30.0%, n=27) sub-group (Table 3.7). Although not analysed because of the very small sample size, similar genotype distributions were identified within the NON sub-group (Table 3.7). No significant differences (p=0.393) were however identified in the *COL12A1* rs970547 genotype frequency distributions between the female larger ACL and CON sub-groups (Table 3.8). Similar genotype distributions were identified within the NON sub-group (Table 3.8). Similar genotype frequency distributions were obtained when control participants with a self-reported history of any previous ligament injury were excluded from the analysis (Appendix C.2). All *COL5A1* and *COL12A1* sub-groups were all in HWE, except for the *COL12A1* rs970547 NON male sub-group (Table 3.7 and 3.8).

Table 3.7. A gender specific comparison of the genotype frequency distributions for *COL5A1* rs12722 (T/C) between control (CON group), anterior cruciate ligament rupture (ACL group) and non-contact mechanism of injury (NON sub-group) groups within the larger South African cohort.

	***COL5A1* rs12722 Genotype**			**n**	**p-value**	**HWE p-value**
	TT	**TC**	**CC**			
Male Participants						
CON	29.1 (41)	53.2 (75)	17.7 (25)	141		0.408
ACL	27.9 (46)	55.8 (92)	16.4 (27)	165	0.899[a]	0.122
NON	30.1 (28)	55.9 (52)	14.0 (13)	93	0.882[b]	0.204
Female Participants						
CON	26.7 (24)	43.3 (39)	30.0 (27)	90		0.211
ACL	44.1 (26)	49.2 (29)	6.8 (4)	59	**0.001**[c]	0.374
NON	38.7 (12)	51.6 (16)	9.7 (3)	31	nd	nd

Values are expressed as percentages with the number of participants indicated in parentheses. The total number (n) of participants genotyped in each individual or combined cohort is also indicated. Values in bold typeset are significant ($p<0.05$). All the ACL ruptures were diagnosed at surgery.

HWE, Hardy-Weinberg Equilibrium; n.d., not determined due to small sample sizes

[a] CON vs. ACL

[b] CON vs. NON

[c] CON vs. ACL and CC vs TC + TT

Table 3.8. A gender specific comparison of the genotype frequency distributions for *COL12A1* (A/G) between control (CON group), anterior cruciate ligament rupture (ACL group) and non-contact mechanism of injury (NON sub-group) groups within the larger SA cohort.

	***COL12A1* rs970547 Genotype**			**n**	**p-value**	**HWE p-value**
	AA	**AG**	**GG**			
Male Participants						
CON	59.6 (84)	36.9 (52)	3.5 (5)	141		0.469
ACL	63.1 (101)	30.6 (49)	6.3 (10)	160	0.345[a]	0.158
NON	64.0 (57)	28.1 (25)	7.9 (7)	89	0.184[b]	**0.036**
Female Participants						
CON	57.6 (49)	41.2 (35)	1.2 (1)	85		0.104
ACL	65.0 (39)	26.7 (16)	8.3 (5)	60	0.393[c]	0.120
NON	62.5 (20)	25.0 (8)	12.5 (4)	32	nd	nd

Values are expressed as percentages with the number of participants indicated in parentheses. The total number (n) of participants genotyped in each individual or combined cohort is also indicated. Values in bold typeset are significant ($p<0.05$). All the ACL ruptures were diagnosed at surgery.

HWE, Hardy-Weinberg Equilibrium

[a] CON vs. ACL

[b] CON vs. NON

[c] CON vs. ACL and AA vs AG + GG

3.3.5 Gene-Gene Interactions and ACL Rupture

When only the males were analysed, no significant associations were identified between any of the four inferred pseudo-haplotypes, constructed from *COL5A1* rs12722 and *COL12A1* rs970547, and risk of ACL rupture (Figure 3.6A). The major T+A inferred pseudo-haplotype was however significantly (p=0.022) over-represented in the female ACL sub-group (50.5%, n=37) when compared to the female CON sub-group (38.1%, n=34) (Figure 3.6B). A similar tendency (p=0.052) was observed for the major T+A inferred pseudo-haplotype to be over-represented in the female NON sub-group (52.4%, n=17) when compared to the female CON sub-group (38.1%, n=34) (Figure 3.6B). The *COL3A1* rs1800255 gene variant was not independently associated with risk of ACL ruptures and only a trend was identified in the male sub-group, therefore analysis of gender-specific gene-gene interactions including *COL3A1* rs180255, *COL5A1* rs12722 and *COL12A1* rs970547 were not determined in this study.

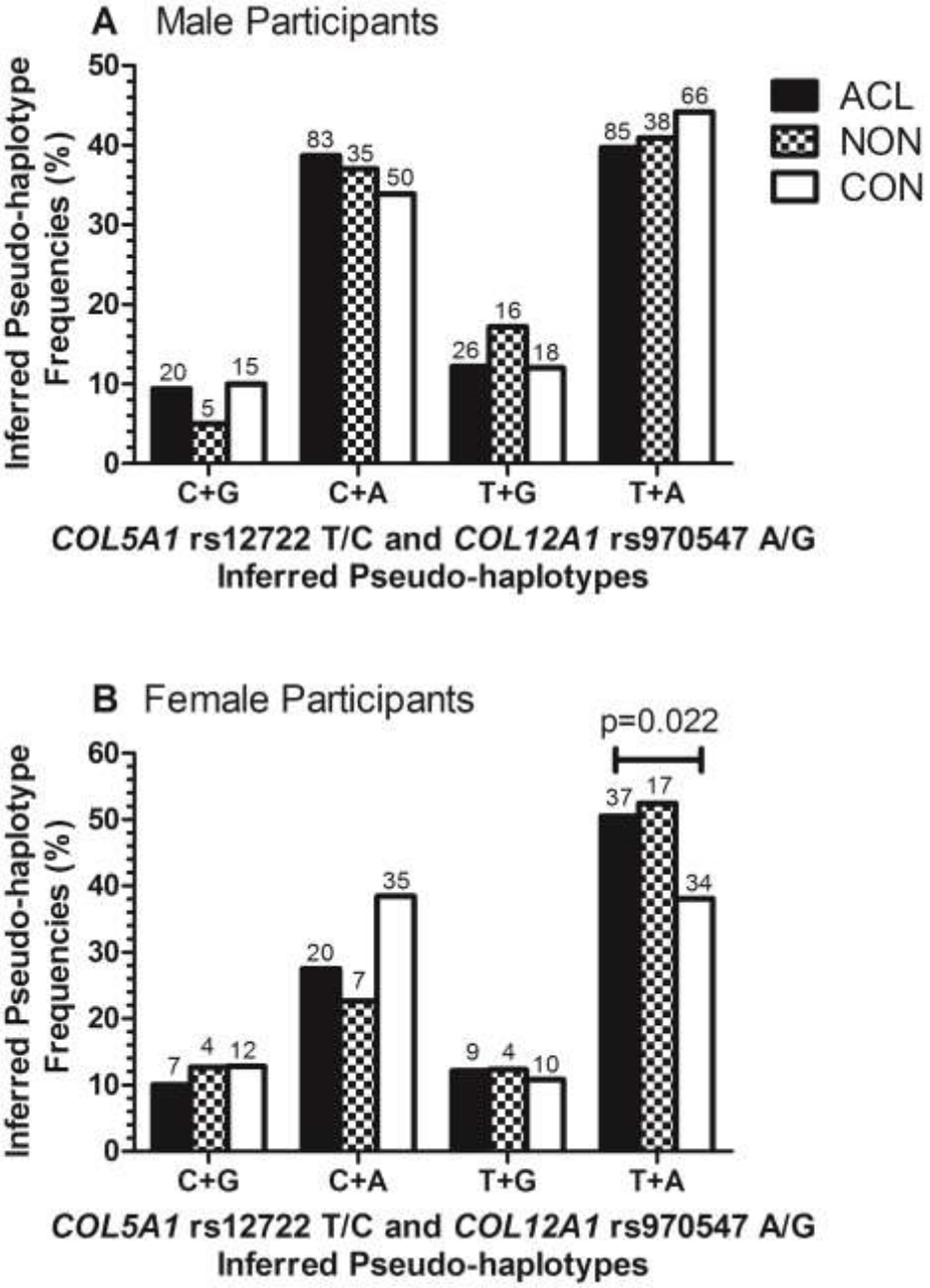


Figure 3.6. Inferred pseudo-haplotype frequency distributions for *COL5A1* rs12722 and *COL12A1* rs970547 in (A) male South African and (B) female South African participants. ACL, Anterior cruciate ligament rupture group. NON, non-contact mechanism of injury sub-group. CON, control group.

3.4 DISCUSSION

The first main finding of this study was that the *COL3A1* rs1800255 gene variant was not independently associated with risk of ACL injuries. Although not significantly associated, a trend was observed for the rare *COL3A1* A allele and AA genotype to be under-represented in the male ACL sub-group when compared to the male CON sub-group. Despite the lack of an independent association, this variant is proposed to be functional [94]. The alanine (rs1800255 G allele) to threonine (rs1800255 A allele) change at position 698 of the COL3A1 peptide is proposed to affect the tensile strength of type III collagen fibres [94]. The hydroxyl side chain of threonine is more hydrophilic than that of alanine and may result in distortion and unwinding of the collagen triple helix of type III collagen [94]. There was also no evidence from this study that this *COL3A1* variant interacted with the *COL5A1* rs12722 and/or *COL12A1* rs970547 variants in modulating the risk of ACL ruptures. In contrast to this finding, a recent study showed that the *COL3A1* rs1800255 AA genotype was associated with increased risk of ACL ruptures in Polish participants (both male and female) [137]. These results suggest that another variant within the *COL3A1* gene or a neighbouring gene may be associated with risk of ACL ruptures. Further studies are required to investigate additional candidates in this region in the aetiology of ACL ruptures.

The second main finding of this study was that *COL6A1* rs35796750 did not associate with risk of ACL rupture in the self-reported Caucasian South African cohort, although it has been associated with a number of other soft-tissue phenotypes [101;138;191;197]. Therefore, this variant was not analysed in the newly

recruited South African subset. The genotype distributions of *COL6A1* rs35796750 were similar to previously reported values for Caucasian populations [197]. This finding does not discount the fact that additional variants within the *COL6A1* gene, or other neighbouring genes, may still play a role in the aetiology of ACL injuries. Further work is required to test this hypothesis.

Since, the *COL5A1* rs12722 (T/C) and *COL12A1* rs970547 (A/G) variants were previously associated with ACL rupture risk in Caucasian South African females [153;154], the gender-specific associations were repeated in a larger South African cohort. Similar to previously published results [153], the *COL5A1* rs12722 CC genotype was significantly associated with reduced risk of ACL ruptures in the larger cohort of female participants. However *COL12A1* rs970547 was not associated with risk of ACL rupture in the larger female cohort. As expected, no significant associations were identified between *COL5A1* rs12722 and *COL12A1* rs970547 and risk of ACL ruptures in the male participants investigated in this study [153;154].

Since the independent analysis of *COL5A1* rs12722 (T/C) and *COL12A1* rs970547 (A/G) within the larger South African cohort was not entirely repeatable, gene-gene interactions were investigated as a possible explanation of this discrepancy. As hypothesised, the major T+A inferred pseudo-haplotype constructed from these variants was associated with increased risk of ACL ruptures in the female participants. When data from an independent Polish cohort were analysed together with the South African cohort from this study the same major T+A inferred pseudo-haplotype was similarly associated with an increased risk of ACL ruptures in the Polish female participants, as well as within the combined female cohorts [137].

These findings are in agreement with the previously reported single association results for these variants [153;154], specifically the T and A alleles of *COL5A1* and *COL12A1*, respectively, have been associated with increased risk of ACL rupture in females. As expected, no significant associations were identified between any of the four inferred *COL5A1* and *COL12A1* pseudo-haplotypes and risk of ACL rupture in the male participants.

In a previous study it was shown that, although not independently associated with risk of ACL rupture, inferred pseudo-haplotypes constructed from variants within the *COL5A1*, *COL11A1* and *COL11A2* genes were associated with risk of Achilles tendinopathy in both a South African and an Australian cohort [71]. This finding, together with the results of this study, highlights the possibility that gene-gene interactions might mask independent single variant associations and that the multigenic nature of ACL ruptures needs to be considered in future research.

As previously discussed (Chapter 2 Section 2.3.2) rs12722 is located within a functional region of the *COL5A1* 3'-UTR. In addition to rs12722, other variants within the 3'-UTR have been reported to be independently associated with chronic Achilles tendinopathy [2;71;107;131]. Future studies should investigate the association of these additional variants with ACL ruptures, especially in females. Although the *COL12A1* rs970547 variant has been proposed to be functional (Chapter 2 Section 2.6.2), it is possible that this variant is tightly linked to another functional variant. Future work is therefore required to examine the association of other *COL12A1* variants with ACL rupture. This work will assist in elucidating possible molecular

mechanisms through which variants within type XII collagen modulate the risk of ACL ruptures.

Finally this study also showed, similarly to Posthumus et al. [152], that South African participants with a family history of any ligament injury were at a 1.8X greater risk of ACL rupture, while a family history of ACL injury increased this to a 9.1X greater risk of ACL rupture. Furthermore, an individual history of any previous ligament and any previous knee ligament injury were also risk factors for ACL rupture in the South African cohort. The ACL group and NON sub-group, and gender sub-groups, also reported playing contact sports for significantly more years than their respective CON groups and sub-groups, while the self-reported hours of non-contact sport were matched between the groups and sub-groups.

In addition, it must be noted that the acute ACL rupture is a complex phenotype with numerous intrinsic and extrinsic risk factors [67]. Therefore, although this study has identified genetic and non-genetic intrinsic risk factors, it is important to acknowledge that they are not wholly responsible for the phenotype but rather contribute to the aetiology that may predispose an individual to risk of acute ACL rupture. This will be discussed in more detail in the final concluding chapter of this thesis (Chapter 10).

The main limitation to this study was the reduced size of the South African sub-groups after gender based analysis. This is not the first study to identify gender-specific associations between gene variants and risk of ACL rupture [153;154], and the reasons for these gender-specific associations remain unknown. As mentioned above, a trend was observed for the rare *COL3A1* rs1800255 A allele and AA

genotype to be under-represented in the male ACL sub-group when compared to the male CON sub-group. A larger cohort of male participants is therefore required to determine if a true independent association exists. In addition, a gene-gene interaction, between *COL5A1* rs12722 and *COL12A1* rs970547, was also significantly associated with risk of ACL ruptures in females only. It is tempting to speculate that these results may be due to direct or indirect interactions between sex hormones and the type V and XII collagen proteins. Interestingly, the expression of matrix metalloproteinase genes, *MMP1* and *MMP3*, is higher in the ACL of women than in men [184]. Furthermore, a gene-gene interaction between *MMP3* and *COL5A1* rs12722 was shown to modulate risk of chronic Achilles tendinopathy [150]. This highlights an indirect mechanism through which these gender-specific associations may be explained, however further studies are required to determine the exact mechanisms involved. A second limitation to this study was that the exposure to extrinsic risk factors for the control participants could not be well documented. Therefore, although care was taken to recruit physically active individuals to match the ACL groups, the level of participation in high risk sports which involve cutting, pivoting and landing could not be accounted for in all control participants. The fact that the male sub-group was not matched for weight is a limitation to the study. It must however be noted that there were no genotype effects on weight and no significant genotype associations were observed for the male sub-group in this study.

In conclusion, the novel main findings of this study are that (i) the *COL3A1* rs1800255 and *COL6A1* rs35796750 gene variants were not associated with risk of ACL injuries in South African participants and (ii) a significant interaction between

the *COL5A1* rs12722 T/C and *COL12A1* rs970547 A/G variants is associated with risk of ACL rupture in females. Specifically the T+A inferred pseudo-haplotype was associated with increased risk of ACL ruptures in female participants. These results highlight the importance of investigating gene-gene interactions in the aetiology of ACL ruptures in multiple independent cohorts. Finally, the results of this chapter are summarised in table 3.9.

Table 3.9. A continuing summary of the results of the chapters in this thesis. The results of chapter 3 are shown.

Investigated Phenotype	Associations	Investigated Gene Variants			
		COL3A1 **rs1800255 (G/A)**	***COL5A1*** **rs12722 (C/T)**	***COL6A1*** **rs35796750 (C/T)**	***COL12A1*** **rs970547 (A/G)**
ACL	**Genotype**	not associated	CC ♀ prev pub[1]	not associated	AA ♀ prev pub[2]
	Haplotype	nd	T ♀[3]	nd	A ♀[3]
Achilles tendinopathy		Chapter 4			
ROM		Chapter 5			
Endurance Performance		Chapter 6			
EAMC		Chapter 7			
Rugby Union		Chapter 8			

Green shading indicates that the allele or genotype is associated with a reduced risk of injury. Red shading indicates that the allele or genotype is associated with an increased risk of injury. Grey shading indicates no independent association. Prev pub, previously published. ♀, associations identified only in females. ACL, anterior cruciate ligament; ROM, range of motion; EAMC, exercise-associated muscle cramps; nd, not determined.

[1] independent association previously reported by Posthumus et al. [153].

[2] independent association previously reported by Posthumus et al. [154], but not idependently associated in this thesis.

[3] TA haplotype also associated with increased risk of female ACL ruptures when a Polish cohort was analysed [137].

CHAPTER 4

GENETIC RISK FACTORS FOR ACHILLES TENDINOPATHY IN TWO INDEPENDENT POPULATIONS

4.1 INTRODUCTION

Similarly to ligaments [59], the collagen fibril is also the major structural component of tendons (Chapter 2 Figure 2.12) [84]. Tendons subjected to large amounts of weight and force may become prone to injury and as a result the majority of tendon injuries affect the Achilles tendon followed by the, patellar, rotator cuff and forearm extensor tendons [162]. It is believed that approximately 30-50% of all sporting injuries occur to tendons and that 6-18% of these occur to the Achilles tendon [128]. Excluding injuries to the surrounding tissues, Achilles tendon injuries can mainly be divided into two separate categories; i) Achilles tendon ruptures and ii) Achilles tendinopathy. Ruptures of the Achilles tendon are acute and occur spontaneously [87], while Achilles tendinopathy is a chronic injury, degenerative in nature and caused by over-use and repetitive straining of the tendon [7], and is the focus of this chapter.

A number of gene variants have been investigated in the aetiology of Achilles tendinopathy [2;50;71;131;155;170;171;179;181], including the candidate variants investigated in this thesis; *COL3A1* rs1800255 (G/A) [170], *COL5A1* rs12722 (C/T)

[131;179] and *COL12A1* rs970547 (A/G) [181]. However, of these three, only *COL5A1* rs12722 was found to be independently significantly associated with Achilles tendinopathy [131;179]. Specifically, the rs12722 CC genotype was significantly over-represented in physically active healthy control participants with no self-reported history of Achilles tendon injury when compared to participants with clinically diagnosed chronic Achilles tendinopathy in independent South African and Australian cohorts [131;179].

As mentioned in the previous chapter, type VI collagen is also present in the solid component of tendons [84], where it is known to interact directly with type V collagen [190]. Furthermore, like type III, V and XII collagen, type VI collagen also facilitates normal collagen fibrillogenesis [129]. The intronic *COL6A1* rs35796750 (T/C) SNP (Chapter 3, Figure 3.4) is located near an intron branch site and is proposed to result in aberrant splicing of the *COL6A1* mRNA [191], which may result in altered or aberrant type VI collagen fibrils. Although *COL6A1* rs35796750 was not associated with ACL ruptures in a South African cohort (Chapter 3), the TT genotype of this variant has previously been associated with increased risk of ossification of the posterior longitudinal ligament [101;191] and the ligamentum flavum [101], as well as diffuse idiopathic skeletal hyperostosis [197].

Since similarities and differences in the genetic associations between ACL and tendon injuries have been reported [131;152-155;179;181], the primary aim of this study was therefore to determine if *COL6A1* rs35796750 is independently associated with risk of chronic Achilles tendinopathy in South African and Australian cohorts. Specifically, it was hypothesised that the *COL6A1* rs35796750 TT genotype would

be associated with increased risk of chronic Achilles tendinopathy. In addition, collagen genes might interact with one another to modulate the risk of musculoskeletal soft tissue injuries (Chapter 3) [71]. The secondary aim was therefore to investigate gene-gene interactions between *COL6A1* rs35796750 and the *COL3A1* rs1800255 and *COL5A1* rs12722 variants, which were previously genotyped in both cohorts, and risk of chronic Achilles tendinopathy in South African and Australian cohorts. It was hypothesised that the *COL5A1* rs12722 T, *COL6A1* rs35796750 T and either of the *COL3A1* rs1800255 alleles may be implicated in gene-gene interactions associated with increased risk of chronic Achilles tendinopathy. The *COL12A1* rs970547 variant was not included in the analysis since it had only been previously genotyped in the South African cohort [181].

4.2 MATERIALS AND METHODS

4.2.1 Participants

Five hundred and seventy-eight unrelated physically active participants, 283 South African (SA) self-reported Caucasian participants and 295 Australian (AUS) self-reported Caucasian participants were recruited and included in this study. The SA cohort included 116 participants with clinically diagnosed chronic Achilles tendinopathy (SA TEN Group) and 167 apparently healthy controls (SA CON Group). The AUS cohort was made up of 85 participants with clinically diagnosed chronic Achilles tendinopathy (AUS TEN Group) and 210 apparently healthy controls (AUS CON Group).

The SA TEN participants were recruited from medical practices within Cape Town in South Africa and were all physically active prior to the onset of the condition. The clinically diagnosis, which was performed as described in Mokone et al (2006) [131], was made and reviewed by experienced sports clinicians. The stringent clinical diagnostic criteria for TEN was a gradual progressive pain over the posterior lower limb in the Achilles tendon area for greater than 6 months. The following six criteria were also used in the diagnosis: (1) early morning pain over the Achilles tendon area, (2) early morning stiffness over the Achilles tendon area, (3) a history of swelling over the Achilles tendon area, (4) tenderness to palpation over the Achilles tendon, (5) palpable nodular thickening over the affected Achilles, or (6) movement of the painful area in the Achilles tendon with plantar-dorsi-flexion (positive ‘‘shift’’ test) [131]. In addition to these criteria, soft-tissue ultrasound examination was performed in a sub-group of subjects to confirm the above diagnosis.

The AUS TEN participants were recruited at the Musculoskeletal Research Centre of La Trobe University, Melbourne, Australia. The clinical diagnosis for the AUS TEN participants was performed as described for the South African participants above and confirmed by soft tissue ultrasound examination

The SA and AUS control subjects were recruited from various recreational sporting clubs. Exclusion criteria included history of current or past fluoroquinolone antibiotic use or previous local corticosteroids injection in the Achilles tendon or the area surrounding the Achilles tendon prior to the onset of symptoms. Furthermore, participants with diagnosed connective tissue disorders or any other systemic

diseases believed to be associated with chronic Achilles tendinopathy, such as, but not limited to, EDS, benign hypermobility joint syndrome, rheumatoid arthritis, systemic lupus erythematosus, hyperparathyroidism, renal insufficiency, diabetes mellitus and familial hypercholesterolemia were also excluded from the study.

Prior to participation in this study, all the participants gave informed written consent (Appendix A.3). In addition, each participant completed personal details, medical history, as well as a sports participation questionnaire (Appendix A.4). This study was approved by the Human Research Ethics Committee of the Faculty of Health Sciences within the University of Cape Town, South Africa and the Human Ethics Committee of La Trobe University, Melbourne, Australia, respectively (Appendix A.1).

4.2.2 Variant Selection

The *COL6A1* rs35796750 variant was chosen for the same reasons as listed in chapter 3 section 3.2.2.

4.2.3 DNA Extraction and Genotyping Methods

For the South African cohort, approximately 4.5ml of venous blood was collected, at the registration of each event, from all participants by venipuncture of a forearm vein into an EDTA vacutainer tube. Blood samples were stored at 4˚C until DNA was extracted, as previously described [108], with minor modifications [131]. An outline of this methodology was described in chapter 3, section 3.2.3. For the Australian cohort

DNA was extracted from approximately 4.5 ml of venous blood using a sequenced extraction technique (FlexiGene DNA Kit, Qiagen P/L, Valencia, California, USA) as per the manufacturer's recommendations. The DNA samples were genotyped at the UCT/MRC Research Unit for Exercise Science and Sports Medicine, University of Cape Town, South Africa for *COL6A1* rs35796750 was performed using custom designed Fluorescence-based Taqman® polymerase chain reaction (PCR) assays (Applied Biosystems, Foster City, CA, USA) as described in chapter 3 section 3.2.3. Investigators were blinded to the phenotypes of the participant samples while genotyping. Two investigators independently confirmed genotyping. Furthermore, a number of positive and negative controls were used to ensure genotyping accuracy. No discrepancies were observed. A total of 404/578 (69.8%) genotypes (200/283 (70.7%) SA genotypes and 204/295 (69.2%) AUS genotypes) were identified for *COL6A1* rs35796750. This low genotyping call rate may be due to degraded DNA samples since recruitment of participants and DNA isolation for these samples was conducted in 2004-2005. In order to avoid genotyping errors, samples that failed twice to amplify during PCR were considered as unsuccessfully genotyped and no further attempts were made to genotype them.

Genotypes, for *COL3A1* rs1800255, *COL5A1* rs12722 and *COL12A1* rs970547, used in the analysis of collagen gene-gene interactions was obtained from previous studies [131;170;179;181]. Genotype and allele frequency distributions and independent association results can be found in appendix C.3-5.

4.2.4 Statistical Analysis

The same statistical tests and haplotype analysis methods were performed as described in chapter 3 section 3.2.4. Significance was accepted when $p<0.05$. No adjustments were made for multiple testing in this study for the same reasons as outlined in chapter 3 section 3.2.4.

4.3 RESULTS

4.3.1 Participant Characteristics

Table 4.1 shows the general characteristics for the South African and Australian participants in this study. The SA TEN group was significantly older at age of injury than the SA CON group at age of recruitment. The SA TEN group was significantly heavier than the SA CON group. The AUS TEN had significantly more male participants when compared to the AUS CON group. All other general characteristics were matched between the SA and AUS TEN and CON groups (Table 4.1). The SA and AUS TEN groups were 6.9 ± 8.1 yrs and 8.9 ± 9.8 yrs older at the time of recruitment than when they were first diagnosed with Achilles tendinopathy. There were no *COL6A1* rs35796750 genotype effects on age (p=0.969), height (p=0.925), weight (p=0.331), BMI (p=0.310) or gender (p=0.400).

Table 4.1. General characteristics for the South African (SA) and Australian (AUS) participants with chronic Achilles tendinopathy (TEN) and apparently healthy controls (CON) recruited for this study.

SA Participants	TEN (116)	CON (167)	p-value
Age (yrs)	41.1 ± 14.1 (99)	36.3 ± 10.9 (160)	**0.001**
Height (cm)	176.4 ± 8.8 (100)	174.8 ± 9.1 (160)	0.540[a]
Weight (kg)	78.0 ± 14.0 (107)	71.9 ± 12.0 (165)	**0.022[b]**
BMI (kg.m^2)	24.8 ± 3.3 (84)	23.5 ± 2.8 (156)	0.087[b]
Gender (% male)	69.3 (79)	62.7 (104)	0.135
AUS Participants	**TEN (85)**	**CON (210)**	**p-value**
Age (yrs)	40.4 ± 14.2 (84)	38.5 ± 12.4 (205)	0.268
Height (cm)	174.0 ± 9.9 (82)	171.5 ± 9.2 (207)	0.057[a]
Weight (kg)	80.5 ± 14.0 (85)	72.7 ± 14.2 (208)	0.240[b]
BMI (kg.m^2)	26.5 ± 3.8 (82)	24.6 ± 3.9 (207)	0.235[b]
Gender (% male)	72.9 (62)	40.2 (84)	**<0.001**

Values are expressed as mean ± standard deviations or as percentages where appropriate. Number of participants is indicated in parentheses. For the TEN groups age, height, weight and BMI are reported from time of injury. For the CON groups age, height, weight and BMI are reported from time of recruitment. Values in bold typeset are significant ($p<0.05$). BMI, body mass index.

[a] Co-varied for gender

[b] Co-varied for age and gender

As mentioned above in section 4.2.1, stringent clinical diagnostic criteria were used during the recruitment of participants by an experienced medical practitioner [131]. The frequency of SA TEN participants clinically diagnosed with tenderness to palpation, early morning stiffness, a history of swelling, early morning pain, palpable thickening, and a positive ‘‘shift’’ test are displayed in table 4.2. In addition, twenty (17.2%) of the SA TEN participants were diagnosed by ultrasound examination of the affected Achilles tendon. Approximately 12.9% (n=15) had a confirmed bilateral Achilles tendinopathy while 8.6% (n=10) reported multiple injuries to the tendon.

Table 4.2. The frequency of the clinical diagnostic criteria within the South African participants with diagnosed chronic Achilles tendinopathy (SA TEN).

Clinical Diagnosis	SA TEN (116)
Tenderness to palpation	21.6 (25)
Early morning stiffness	16.4 (19)
History of swelling	12.9 (15)
Early morning pain	10.3 (12)
Palpable thickening	7.8 (9)
Positive ‘‘shift’’ test	7.8 (9)
At least one of the above	38.8 (45)
All of the above	5.2 (6)

Values are expressed as percentages. Number of participants is indicated in parentheses.

All participants in the AUS TEN group were examined using ultrasound examination of the affected Achilles tendon to confirm pathology in addition to their clinical diagnosis [179]. Furthermore 51.8% (n=44) of the AUS TEN participants were diagnosed with bilateral chronic Achilles tendinopathy while 27.1% (n=23) reported multiple injuries to the tendon.

Running was found to be the predominant sporting activity resulting in injury (19.0%, n=22) in the SA cohort. The SA cohort was matched for the mean number of years participating in running (SA CON 8.7 ± 9.5 yrs, n=57, SA TEN 8.3 ± 10.7 yrs, n=41, p=0.821). However, the SA CON group (3.4 ± 3.1 hrs, n=57) reported significantly (p=0.017) more hours of running training when compared to the SA TEN group (1.9

± 2.4 hrs, n=36). Although all AUS participants were physically active individuals, the type of sporting activities performed, the hours of training and the frequency of activity were not recorded.

4.3.2 COL6A1 rs35796750 and Achilles Tendinopathy

The individual genotype frequency distributions for *COL6A1* rs35796750 within the South African and Australian cohorts, as well as in all participants, are provided in table 4.3. There were no significant differences in the *COL6A1* rs35796750 genotype or allele frequency distributions between either cohort, or when they were combined (Tables 4.3). Since no previous studies have shown a gender-specific association with the *COL6A1* rs35796750 variant, gender-specific analysis of risk of Achilles tendinopathy was not investigated. The SA TEN (p=0.806) and CON (p=0.565) groups and the AUS TEN (p=0.188) group were in HWE (Table 4.3). The AUS CON group was out of HWE (p=0.011) (Table 4.3). In addition, when the cohorts were combined, both the TEN (p=0.043) and CON (p=0.023) groups were out of HWE (Table 4.3).

Table 4.3. A comparison of the genotype and allele frequency distributions for *COL6A1* rs35796750 (T/C) between the South African and Australian Achilles tendinopathy groups and their respective control groups. Genotype and allele frequency distributions after combining the cohorts are also reported.

	***COL6A1* rs35796750 Genotype**			**n**	**Genotype**	**HWE**	**Minor**	**Allele**
	TT	**TC**	**CC**		**p-value**		**Allele**	**p-value**
SA TEN	40.0 (32)	48.8 (39)	11.3 (9)	80		0.806	35.6 (57)	
SA CON	37.5 (45)	50.0 (60)	12.5 (15)	120	0.925	0.565	37.5 (90)	0.703
AUS TEN	27.9 (17)	59.0 (36)	13.1 (8)	61		0.188	42.6 (52)	
AUS CON	26.6 (38)	60.1 (86)	13.3 (19)	143	0.982	**0.011**	43.4 (124)	0.913
All TEN	34.8 (49)	53.2 (75)	12.1 (17)	141		**0.043**	38.7 (109)	
All CON	31.6 (83)	55.5 (146)	12.9 (34)	263	0.806	**0.023**	40.7 (214)	0.574

Values are expressed as percentages with the number of participants indicated in parentheses. The total number (n) of participants genotyped in each individual or combined cohort is also indicated. Significant p-values are indicated in bold.
HWE, Hardy-Weinberg Equilibrium; n.d., not determined due to small sample sizes; SA, South African; AUS, Australia; All, South Africa and Australia; TEN, participants with diagnosed Achilles tendinopathy; CON, apparently healthy control participants.

4.3.3 Gene-gene Interactions and Achilles Tendinopathy

Gene-gene interactions between *COL6A1* rs35796750 and the previously genotyped *COL3A1* rs1800255 (G/A) [170] and *COL5A1* rs12722 (T/C) [131;179] were investigated. The South African and Australian cohorts were combined to increase the sample size for the analysis, since the genotype frequency distributions were

similar between the TEN and CON groups for all three variants. The *COL12A1* rs970547 variant was therefore not included in these gene-gene interaction analyses since it has not been genotyped in the Australian cohort (Appendix C.6) [181]. The A+C+C inferred pseudo-haplotype, constructed from *COL3A1* rs1800255, *COL5A1* rs12722 and *COL6A1* rs35796750, was significantly (p=0.005) under-represented in the TEN group (0.0%, n=0) when compared to the CON group (7.6%, n=28) (Figure 4.1). Interestingly, the A+C inferred pseudo-haplotype, constructed from *COL3A1* rs1800255 and *COL6A1* rs35796750, was also significantly (p=0.038) under-represented in the TEN group (5.0%, n=10) when compared to the CON group (11.5%, n=43) (Appendix C.7). All the two-gene inferred pseudo-haplotypes are presented in appendix C.7.

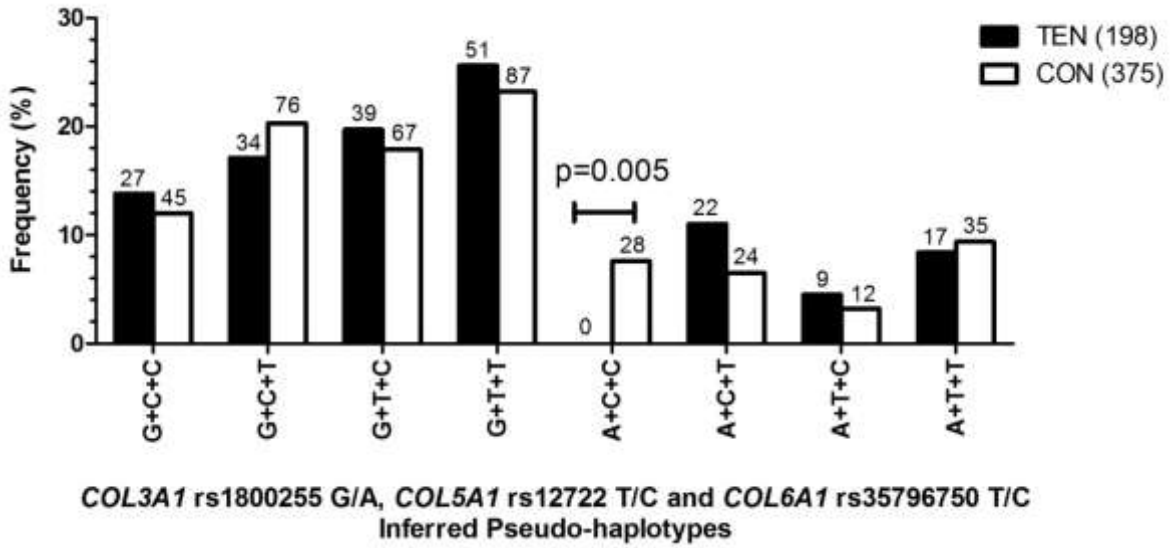


Figure 4.1. Inferred pseudo-haplotype frequencies, constructed from *COL3A1* rs1800255, *COL5A1* rs12722 and *COL6A1* rs37596750, in South African and Australian participants with clinically diagnosed Achilles tendinopathy (TEN group) and apparently healthy controls (CON group). Numbers of participants are listed above each column.

4.4 DISCUSSION

The *COL6A1* rs35796750 variant was not independently associated with risk of chronic Achilles tendinopathy in a South African or Australian cohort. As mentioned previously this variant was selected because of proposed functionality [191], previous association studies with other connective tissue phenotypes [101;191;197], as well as the role of type VI collagen in fibrillogenesis [129] and its interaction with type V collagen [190]. Therefore, although *COL6A1* rs35796750 was not associated with risk of Achilles tendinopathy in this study it does not exclude the fact that additional *COL6A1* variants may still play a role in in the aetiology of Achilles tendinopathy.

Gene-gene interactions between *COL6A1* rs35796750 and the previously investigated *COL3A1* rs1800255 [170] and *COL5A1* rs12722 [131;179] variants were also explored. A significant two-gene inferred pseudo-haplotype constructed from *COL3A1* rs1800255 and *COL6A1* rs35796750 was associated with risk of Achilles tendinopathy. Consistent with the hypothesis, the A+C inferred pseudo-haplotype was associated with reduced risk of Achilles tendinopathy. Similarly, the A+C+C inferred pseudo-haplotype, constructed from *COL3A1* rs1800255, *COL5A1* rs12722 and *COL6A1* rs35796750, was also associated with reduced risk of ACL ruptures. These results implicate, for the first time, the *COL3A1* rs1800255 and *COL6A1* rs35796750 gene variants as risk factors for Achilles tendinopathy. Further studies are required to investigate how these, and other, collagen types may interact to modulate risk of Achilles tendinopathy.

In addition, it must be noted that chronic Achilles tendinopathy is a complex phenotype with numerous intrinsic and extrinsic risk factors [128]. Therefore, although this study has identified novel genetic risk factors, it is important to acknowledge that they are not wholly responsible for the phenotype but rather contribute to the aetiology that may predispose an individual to risk of chronic Achilles tendinopathy. This will be discussed in more detail in the final concluding chapter of this thesis (Chapter 10).

The main limitation to this study was low genotyping call rate for the *COL6A1* rs35796750 variant. A possible reason for this is described in the genotyping section 4.2.3. These low call rates may also be responsible for the AUS CON group being out of HWE. The differences in participant characteristics between the SA TEN and CON and AUS TEN and CON groups are a further limitation to this study. Both the SA and AUS TEN groups weighed more than their corresponding SA and AUS CON groups however participants were not recruited at the time of the injury and this data was self-reported. The TEN participants from both cohorts anecdotally reported an increase in weight after injury however this cannot be confirmed. Finally, the lack of physical activity data, including the type of sporting activities performed, the hours of training and the frequency of activity, for the Australian cohort is also a limitation to this study.

In conclusion, the *COL6A1* rs35796750 variant is not independently associated with risk of Achilles tendinopathy. However, the novel finding of this study was that the *COL3A1* rs1800255 and *COL6A1* rs35796750 variants interact to modulate risk of Achilles tendinopathy in the combined South African and Australian cohorts. In

addition, the *COL3A1* rs1800255, *COL5A1* rs12722 and *COL6A1* rs35796750 further modulate the risk of Achilles tendinopathy. Future studies are required to investigate how these, and other, collagen types may interact to modulate risk of chronic Achilles tendinopathy. The results of this chapter are summarised in table 4.

Table 4.4. A continuing summary of the results of the chapters in this thesis. The results of chapter 4, and all preceding chapters, are shown.

Investigated Phenotype	Associations	Investigated Gene Variants			
		COL3A1 **rs1800255 (G/A)**	***COL5A1*** **rs12722 (C/T)**	***COL6A1*** **rs35796750 (C/T)**	***COL12A1*** **rs970547 (A/G)**
ACL	**Genotype**	not associated	CC ♀ prev pub[1]	not associated	AA ♀ prev pub[2]
	Haplotype	nd	T ♀[3]	nd	A ♀[3]
Achilles tendinopathy	**Genotype**	not associated prev pub[4]	CC prev pub[5]	not associated	not associated prev pub[6]
	Haplotype	A	C	C	nd
ROM		Chapter 5			
Endurance Performance		Chapter 6			
EAMC		Chapter 7			
Rugby Union		Chapter 8			

Green shading indicates that the allele or genotype is associated with a reduced risk of injury. Red shading indicates that the allele or genotype is associated with an increased risk of injury. Grey shading indicates no independent association. Prev pub, previously published. ♀, associations identified only in females. ACL, anterior cruciate ligament; ROM, range of motion; EAMC, exercise-associated muscle cramps; nd, not determined.

[1] independent association previously reported by Posthumus et al. [153].
[2] independent association previously reported by Posthumus et al. [154], but not independently associated in this thesis.
[3] T+A pseudo-haplotype also associated with increased risk of female ACL ruptures when a Polish cohort was analysed [137].
[4] no independent association previously reported by Saunders [170]
[5] independent association previously reported by Mokone et al. [131] and repeated by September et al. [179].
[6] no independent association previously reported by September et al. [181].

CHAPTER 5

GENETIC MARKERS FOR RANGE OF MOTION

The data presented in this chapter has been published in a condensed form in the peer-reviewed article: O'Connell K, Posthumus M, Collins M. No association between *COL3A1*, *COL6A1* or *COL12A1* gene variants and range of motion. *Journal of Sports Sciences.* 2013;31(2):181-7.

5.1 INTRODUCTION

The causes of anterior cruciate ligament (ACL) ruptures, chronic Achilles tendinopathy, as well as, other musculoskeletal soft tissue injuries are poorly understood and complex [67;128]. As reported in the previous chapters multiple extrinsic and intrinsic risk factors, including a genetic component, have been associated with these injuries [71;131;137;150;152-155;160;171;179-181]. Joint range of motion (ROM) or flexibility has also been proposed as one of the intrinsic risk factors for ACL ruptures and chronic Achilles tendinopathy [95;96;103].

Joint range of motion may be defined as the range of motion available to a joint or group of joints, that does not result in injury, and is influenced by the muscles, tendons, ligaments and bones associated with that joint [6]. Like musculoskeletal soft tissue injuries, joint ROM is also a complex multifactorial phenotype [6;11;70;119;204]. A number of intrinsic factors, which includes a heritable component [11;119], and extrinsic factors are known to influence inter-individual

range of motion [6;70;204]. It has been estimated, in classical twin studies, that approximately 64-70% of the variability in normal (non-pathological) joint range of motion may be heritable [11;119]. *COL5A1* rs12722 (C/T), which is a focus of this thesis, was the first variant reported to be associated with normal joint range of motion in independent cohorts [1;28;29;40]. Specifically, the CC genotype has been reported to be associated with an age-dependent increase in sit-and-reach ROM in apparently healthy physically active individuals [29]. In addition, analysis of ultra-endurance athletes, who participated in a 56km road race, showed that the C and T alleles were over-represented in the “flexible-fast” and “inflexible-slow” groups respectively [28]. In contrast to these findings, the CT genotype was also shown to be associated with both increased knee hyperextension and general joint laxity in females only [13] and with increased sit-and-reach and straight leg raise ROM in uninjured participants as well as those with Achilles tendon injuries [40], while the CT genotype was also significantly over-represented in a cohort of Italian high-level international rhythmic gymnasts when compared to an active control group [196]. As expected, athletes with a -/- *COL5A1* rs71746744 (-/AGGG) genotype also had on average a significantly higher pre-race sit-and-reach ROM [1]. As previously mentioned (Chapter 2 Figure 2.7) both rs12722 and rs71746744 are located within the function 3’-untranslated region (UTR) of the *COL5A1* gene [107].

A genetic continuum is proposed for collagen genes, such as *COL5A1* (Chapter 2 Section 2.2), in which rare mutations result in severe disorders. Extreme joint hypermobility, which is one of the clinical features of the inherited connective tissue disorders such as classical (types 1 and 2) and hypermobility (type 3) types of Ehlers-Danlos Syndrome (EDS), are caused by rare mutations within collagen genes

[19;33;122-124;186;218]. Approximately a half of the classical type of EDS is caused by mutations within *COL5A1* [19;123]. These mutations within the *COL5A1* gene result in a 50% reduction in type V collagen production, aberrant fibrillogenesis and altered mechanical properties of connective tissues [19;123]. *COL3A1* mutations cause the vascular type of EDS (type 4), which has been reported to cause aberrant fibrillogenesis [186]. As previously discussed (Chapter 2 Section 2.6), in addition to types V and III collagen, types VI and XII collagen, as well as other non-collagen proteins, are all implicated in fibrillogenesis [56;115;129;141;213]. Recently mutations within the *COL12A1* gene have also been shown to cause a novel EDS [218], while the *COL12A1* rs970547 AA genotype was also associated with increased knee hyperextension in females [13]. Mutations within the *COL6A1* gene do not result in any of the EDS types, however patients with EDS resulting from tenascin-X deficiency show reduced *COL6A1* expression [31]. Furthermore, mutations in *COL6A1* do cause Bethlem myopathy [17;109] and Ullrich congenital muscular dystrophy [17;109;140], of which a symptom is joint hypermobility.

In summary (1) Rare mutations within the four collagen genes investigated in this thesis (*COL3A1*, *COL5A1*, *COL6A1* and *COL12A1*) cause joint hypermobility, (2) joint ROM is an intrinsic risk factors for ACL ruptures and chronic Achilles tendinopathy, (3) the collagens encoded by these candidate genes are all involved in fibrillogenesis, (4) all four collagen gene common variants investigated in this study are either independently associated or form part of inferred pseudo-haplotypes associated with modifying the risk ACL ruptures and/or chronic Achilles tendinopathy. Since these observations are in agreement with the genetic continuum hypothesis presented in chapter 2, the primary aim of this study in the thesis was to

investigate the independent association of the *COL3A1* rs1800255 (G/A), *COL6A1* rs35796750 (T/C) and *COL12A1* rs970547 (A/G) variants with three measures of range of motion, namely sit-and-reach, straight leg raise (SLR) and total shoulder rotation (ShTR) range of motion. The secondary aim of this study was to determine if any gene-gene interactions between these collagen genes were associated with the sit-and-reach ROM.

5.2 MATERIALS AND METHODS

5.2.1 Participants

Three hundred and fifty apparently healthy, injury free and physically active Caucasian participants were recruited from fitness centers/clubs and two road running events within the greater Cape Town area of South Africa. The inclusion criteria have been previously described [28;29]. Briefly, only participants that (1) were non-obese (Body mass index (BMI) less than 30 $kg.m^{-2}$), (2) were over the age of 18 years and (3) reported no serious injuries in the last 12 months were recruited.

All participants gave written informed consent (Appendix A.3) and completed a physical activity, medical history and flexibility training questionnaire (Appendix A.4). Ethical approval for the study was obtained from the Human Research Ethics Committee of the University of Cape Town (Appendix A.1). In addition, permission to conduct this study was also granted by the organising committee(s) of the respective running races, where applicable.

5.2.2 Anthropometric and Range of Motion Measurements

Anthropometric measurements and range of motion assessments were measured as previously described [28;29]. Briefly, three clinically validated ROM assessments were performed on the subjects at two separate testing sessions. Two lower body assessments, the sit-and-reach test [212] and straight leg raise (SLR) test [63], were performed at separate visits to prevent one influencing the other. Participants were requested to return for their second testing session between one and ten days after their initial session. The third assessment, an upper body assessment, was the shoulder internal (IR) and external rotation (ER) in 90° abduction [10;24]. This assessment was always performed on the same day as the SLR test.

Both the sit-and-reach and straight leg raise (SLR) range of motion tests are indirect measures of lumbar and hamstring musculotendinous unit range of motion [63;212]. Total shoulder rotation (ShTR) was calculated by combining measures of internal and external rotation at 90^{o} of abduction [10;24]. Measurements of straight leg raise and total shoulder rotation range of motion were determined for dominant and non-dominant limbs. Measures of sit-and-reach, straight leg raise and total shoulder rotation range of motion were recorded in triplicate and the best score used.

5.2.3 Variant Selection

Three independent sequence variants were investigated in this study. Namely *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547. The reasons for their selection have been stated in chapter 3, section 3.2.2. Furthermore, these variants have been independently associated with, or interact to modulate, the risk of musculoskeletal soft tissue injuries (Chapter 3 and 4).

5.2.4 DNA Extraction and Genotyping Methods

Approximately 4.5ml of venous blood was collected from all participants by venipuncture of a forearm vein into an EDTA vacutainer tube. These samples were stored at 4˚C until DNA extraction, as previously described [108], with minor modifications [131]. An outline of this methodology was described in chapter 3, section 3.2.3.

Participants were genotyped for *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547. Investigators were blinded to the performance results of the participant samples while genotyping. Two investigators independently confirmed genotyping. A number of positive and negative controls were used to ensure quality and accuracy of genotyping. No discrepancies were observed. Of the 350 participants recruited only 314 (89.7%), 313 (89.4%) and 315 (90.0%) genotypes were obtained for *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 respectively.

Genotyping of *COL3A1* rs1800255 and *COL6A1* rs35796750 was performed using custom designed Fluorescence-based Taqman® polymerase chain reaction (PCR) assays (Applied Biosystems, Foster City, CA, USA) as described in chapter 3. Additional genotyping information is available in appendix B.

COL12A1 rs970547 was genotyped as previously described [39;42]. Briefly, fragments containing the *COL12A1 Alu*I RFLP were amplified by polymerase chain reaction (PCR). The polymerase chain reaction products were then digested with *Alu*I to produce 599bp and 16bp fragments for the G allele; 460, 139 and 16bp fragments for the A allele. The digest fragments were separated, together with a 100bp DNA ladder, on 6% non-denaturing polyacrylamide electrophoresis gels and visualized by SYBER Gold staining (*Invitrogen Molecular Probes*TM, Oregon, USA). The gels were photographed under UV light using a Uvitec photodocumentation system (Uvitec Limited, Cambridge, UK) and the sizes of the DNA fragments determined. Additional genotyping information is available in appendix B. All genotyping was done at the UCT/MRC Research Unit for Exercise Science and Sports Medicine, University of Cape Town, Cape Town, South Africa.

5.2.5 Statistical Analysis

The same statistical tests and haplotype analysis methods were performed as described in chapter 3 section 3.2.4. Although not significantly different, large differences (≥10%) were identified between the frequencies of male participants in the SNP genotype groups. In addition, gender is a known factor in the aetiology of ROM [6;73]. Therefore, all of the analyses presented in this study are presented for

males and females separately or, when the whole cohort was analysed, were co-varied for gender. Furthermore, analysis of *COL6A1* rs35796750 was also co-varied for weight due to the significant inter-genotype differences identified. Bivariate correlations were used to investigate the relationship between participant characteristics and measures of range of motion. Significance was accepted when $p<0.05$. No adjustments were made for multiple testing in this study for the same reasons as outlined in chapter 3 section 3.2.4.

5.3 RESULTS

5.3.1 Participant Characteristics

Average age, height, weight, BMI and ROM measurements of all participants, as well as the male and female participants, are presented in table 5.1. There were no significant differences in the average ages of the male and female participants. As expected the males were significantly taller and heavier than their female counterparts. On average the female participants had significantly higher ($p<0.001$) measures of range of motion than male participants for all measurements (Table 5.1), except for non-dominant shoulder total rotation which only showed a tendency to be higher in females after adjusting for height, weight and self-reported flexibility training. Mean time spent on flexibility training was 1.3 ± 2.7 minutes/week (n=112).

Correlations between participant age, weight, height, BMI and flexibility training and all measures of range of motion are presented in table 5.2. A significant negative correlation was identified between age and dominant SLR ROM in males only, while

a significant positive correlation was identified between BMI and non-dominant ShTR ROM in females only (Table 5.2). No other significant correlations were identified.

Table 5.1. General characteristics, as well as upper and lower body range of motion (ROM) measurements for all participants. A comparison between male and female participants is also presented.

	All (350)	Participant Sex		p value [a]	p value [b]
		Male (216)	Female (134)		
Age (yrs)	32.8 ± 11.8 (350)	33.2 ± 12.3 (216)	32.4 ± 10.8 (134)	0.570	nd
Height (cm)	174.4 ± 9.4 (310)	180.0 ± 6.7 (187)	166.1 ± 6.4 (122)	**<0.001**	nd
Weight (kg)	72.6 ± 13.0 (310)	79.3 ± 10.9 (187)	62.5 ± 8.6 (122)	**<0.001**	nd
BMI ($kg.m^{-2}$)	23.8 ± 3.9 (308)	24.4 ± 2.8 (187)	23.0 ± 5.0 (122)	**0.001**	nd
SR ROM (mm)	270.1 ± 110.2 (344)	245.8 ± 106.9 (213)	309.8 ± 104.1 (131)	**<0.001**	**0.009**
Non-Dom SLR (°)	87.8 ± 20.4 (162)	81.1 ± 17.8 (95)	97.4 ± 20.1 (67)	**<0.001**	**<0.001**
Dom SLR (°)	92.5 ± 20.4 (162)	84.8 ± 17.6 (95)	103.4 ± 19.1 (67)	**<0.001**	**0.001**
Non-Dom ShTR (°)	204.2 ± 28.6 (157)	196.7 ± 25.0 (93)	215.2 ± 30.2 (64)	**<0.001**	0.056
Dom ShTR (°)	198.7 ± 26.6 (159)	190.5 ± 23.3 (94)	210.4 ± 26.7 (65)	**<0.001**	**<0.001**

Values are expressed as mean ± standard deviation. Number of participants with non-missing data is indicated in parentheses. Values in bold typeset are significant ($p<0.05$). BMI, body mass index; SR, sit-and-reach; mm, millimetres; Non-Dom, non-dominant; SLR, straight leg raise; Dom, dominant; ShTR, shoulder total rotation; nd, not determined.

[a] male vs female

[b] co-varied for height, weight and flexibility training

Table 5.2. Correlations between Participant General Characteristics and Measures of Range of Motion (ROM) in males and females.

Male Participants

	SR ROM (mm)		**Non-Dom SLR** (°)		**Dom SLR** (°)		**Non-Dom** ShTR (°)		**Dom ShTR** (°)	
Age (yrs)	r=-0.119 p=0.085	(212)	r=-0.161 p=0.119	(95)	r=-0.216 **p=0.035**	(95)	r=-0.087 p=0.408	(93)	r=0.182 p=0.862	(94)
Weight (kg)	r=-0.037 p=0.620	(184)	r=0.124 p=0.234	(94)	r=0.007 p=0.944	(94)	r=-0.089 p=0.398	(92)	r=-0.020 p=0.848	(93)
Height (m)	r=0.046 p=0.537	(184)	r=0.057 p=0.588	(94)	r=-0.086 p=0.408	(94)	r=-0.097 p=0.358	(92)	r=0.093 p=0.374	(93)
BMI ($kg.m^{-2}$)	r=-0.085 p=0.254	(182)	r=0.117 p=0.267	(92)	r=0.063 p=0.552	(92)	r=-0.060 p=0.574	(90)	r=-0.073 p=0.492	(91)
Flexibility Training (min/wk) [a]	r=0.050 p=0.553	(144)	r=0.179 p=0.089	(91)	r=0.134 p=0.207	(91)	r=-0.022 p=0.839	(89)	r=0.160 p=0.131	(90)

Female Participants

	SR ROM (mm)		**Non-Dom SLR** (°)		**Dom SLR** (°)		**Non-Dom** ShTR (°)		**Dom ShTR** (°)	
Age (yrs)	r=0.022 p=0.804	(131)	r=-0.070 p=0.569	(67)	r=-0.018 p=0.884	(67)	r=-0.150 p=0.236	(64)	r=-0.209 p=0.096	(65)
Weight (kg)	r=0.049 p=0.596	(119)	r=0.041 p=0.744	(67)	r=-0.043 p=0.729	(67)	r=0.135 p=0.289	(64)	r=0.134 p=0.289	(65)
Height (m)	r=-0.040 p=0.668	(119)	r=-0.092 p=0.460	(67)	r=-0.114 p=0.357	(67)	r=-0.152 p=0.231	(64)	r=0.004 p=0.974	(65)
BMI ($kg.m^{-2}$)	r=0.108 p=0.241	(119)	r=0.117 p=0.347	(67)	r=0.032 p=0.798	(67)	r=0.255 **p=0.042**	(64)	r=0.159 p=0.205	(65)
Flexibility Training (min/wk) [a]	r=0.115 p=0.308	(81)	r=0.043 p=0.737	(64)	r=0.054 p=0.669	(64)	r=-0.102 p=0.434	(61)	r=-0.090 p=0.487	(62)

Number of participants with non-missing data is indicated in parentheses. Values in bold typeset are significant ($p<0.05$).

[a] There were also no significant correlations of flexibility training with any of the ROM measurements when only those participants with a self-reported history of flexibility training were included in the analysis.

SR, sit-and-reach; mm, millimetres; Non-Dom, non-dominant; SLR, straight leg raise; Dom, dominant; ShTR, shoulder total rotation; BMI, body mass index; yrs, years, kg, kilogram; m, metres; min, minutes; wk, week

Participants with a self-reported history of flexibility training (47.7%, n=112) had significantly higher (p=0.007) measures for sit-and-reach range of motion measurements than the 123 participants with no self-reported history of flexibility training (Table 5.3). No significant differences in the other measures of range of motion were identified when participants with a self-reported history of flexibility training were compared to those with no history of flexibility training (Table 5.3).

No significant genotype interactions were identified between *COL3A1* rs1800255, *COL6A1* rs35796750 or *COL12A1* rs970547 and participant age, gender, height, BMI, minutes of flexibility training per week or hours of exercise per week (Table 5.4). A significant genotype effect was determined between only *COL6A1* rs35796750 and participant weight (TT: 73.9 kg, TC: 70.7 kg, CC: 75.7 kg, p=0.035).

Table 5.3. Differences in range of motion (ROM) measurements between participants with a self-reported history (Yes) and no history (No) of weekly flexibility training.

	Flexibility Training		p value [b]
	Yes (112) [a]	No (123)	
SR ROM (mm)	288.4 ± 113.7 (107)	239.7 ± 108.0 (121)	0.007
Non-Dom SLR (°)	91.7 ± 21.4 (70)	83.6 ± 18.8 (85)	0.066
Dom SLR (°)	96.5 ± 20.7 (70)	88.1 ± 19.2 (85)	0.066
Non-Dom ShTR (°)	204.9 ± 28.2 (68) [c]	202.8 ± 29.3 (82)	0.781
Dom ShTR (°)	200.3 ± 27.3 (68) [c]	196.1 ± 26.0 (84)	0.871

Values are expressed as mean ± standard deviation. Number of participants is indicated in parentheses. Values in bold typeset are significant ($p<0.05$).

SR, sit-and-reach; mm, millimetres; Non-Dom, non-dominant; SLR, straight leg raise; Dom, dominant; ShTR, shoulder total rotation

[a] All 112 participants reported doing lower body flexibility (111 hamstrings, 106 quadriceps, 80 soleus, 95 gastrocnemius, 74 groin) training.

[b] p values adjusted for gender.

[c] Non-Dom ShTR and Dom ShTR for participants who reported Upper body Flexibility training was 206.5 ± 28.7° (43) and 199.3 ± 24.2° (43) respectively.

Table 5.4. Genotype effects of *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 on physiological characteristics of participants.

Polymorphism	p-values						
	Age	Height	Weight	BMI	Gender	Flex training (min/wk)	Flex training (hr/wk)
***COL3A1* rs1800255**	0.126	0.725	0.668	0.274	0.513	0.130	0.413
***COL6A1* rs35796750**	0.108	0.066	**0.035**	0.743	0.136	0.359	0.684
***COL12A1* rs970547**	0.977	0.232	0.788	0.596	0.197	0.863	0.108

BMI, body mass index. Flex, flexibility. p-values in bold typeset indicate significant differences ($p<0.05$).

5.3.2 Collagen Genes and Range of Motion

The *COL3A1* rs1800255 (p=0.180), *COL6A1* rs35796750 (p=0.219) and *COL12A1* rs970547 (p=0.988) variants were not significantly associated with sit-and-reach range of motion (Table 5.5). In addition, no significant associations were identified between dominant straight leg raise, non-dominant straight leg raise, dominant total shoulder rotation or non-dominant total shoulder rotation range of motion and *COL3A1* rs1800255, *COL6A1* rs35796750 or *COL12A1* rs970547 (Table 5.5). In addition, no significant genotype interactions were identified for any of the measures of range of motion after further adjusting for self-reported history of flexibility training (Table 5.5).

Since the *COL5A1* rs12722 gene variant was previously shown to be associated with sit-and-reach ROM in older participants (≥35yrs old) [29], associations between *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 were investigated in an older group (≥35yrs old) (Table 5.6). The *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 were not significantly associated with sit-and-reach range of motion in this older group (Table 5.6). Analysis of the other measures of ROM could not be investigated due to small sample sizes in the older group (Table 5.6). Furthermore, no significant age-genotype interaction effects were identified between any of the range of motion measurements and *COL3A1* rs1800255 ($p \geq 0.143$), *COL6A1* rs35796750 ($p \geq 0.104$) or *COL12A1* rs970547 ($p \geq 0.100$) (Table 5.7).

Table 5.5. Differences in range of motion (ROM) measurements between the *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 genotype groups.

	***COL3A1* rs1800255 Genotypes**			**p value [a]**	**p value [c]**
	GG (160)	**GA (124)**	**AA (25)**		
SR ROM (mm)	279 ± 110 (160)	268 ± 106 (124)	243 ± 122 (25)	0.180	0.090
Non-Dom SLR (°)	90 ± 21 (75)	87 ± 20 (57)	83 ± 20 (8)	0.423	0.335
Dom SLR (°)	94 ± 21 (75)	91 ± 19 (57)	89 ± 21 (8)	0.389	0.265
Non-Dom ShTR (°)	207 ± 32 (72)	205 ± 25 (55)	185 ± 32 (8)	0.120	0.114
Dom ShTR (°)	200 ± 30 (73)	200 ± 24 (56)	193 ± 24 (8)	0.773	0.715
	***COL6A1* rs35796750 Genotypes**			**p value [b]**	**p value [d]**
	TT (115)	**TC (139)**	**CC (54)**		
SR ROM (mm)	272 ± 103 (115)	262 ± 118 (139)	284 ± 111 (54)	0.219	0.426
Non-Dom SLR (°)	89 ± 21 (48)	84 ± 22 (67)	93 ± 18 (24)	0.128	0.178
Dom SLR (°)	94 ± 24 (48)	89 ± 19 (67)	97 ± 16 (24)	0.051	0.078
Non-Dom ShTR (°)	204 ± 28 (46)	200 ± 30 (64)	206 ± 34 (24)	0.687	0.686
Dom ShTR (°)	197 ± 25 (47)	195 ± 28 (65)	206 ± 32 (24)	0.293	0.209
	***COL12A1* rs970547 Genotypes**			**p value [a]**	**p value [c]**
	AA (187)	**AG (112)**	**GG (11)**		
SR ROM (mm)	271 ± 107 (187)	265 ± 119 (112)	266 ± 102 (11)	0.988	0.877
Non-Dom SLR (°)	91 ± 20 (95)	85 ± 21 (57)	74 ± 12 (5)	0.059	0.077
Dom SLR (°)	96 ± 22 (95)	89 ± 118 (57)	84 ± 8 (5)	0.093	0.125
Non-Dom ShTR (°)	209 ± 30 (93)	197 ± 25 (54)	206 ± 32 (5)	0.144	0.208
Dom ShTR (°)	202 ± 29 (93)	193 ± 23 (56)	202 ± 20 (5)	0.214	0.241

Values are expressed as mean ± standard deviation. Number of participants is indicated in parentheses. SR, sit-and-reach; mm, millimetres; Non-Dom, non-dominant; SLR, straight leg raise; Dom, dominant; ShTR, shoulder total rotation.

[a] p values adjusted for gender.

[b] p values adjusted for gender and weight.

[c] p values adjusted for gender and self-reported history of flexibility training.

[d] p values adjusted for gender, weight and self-reported history of flexibility training.

Table 5.6. Differences in range of motion (ROM) measurements between the *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 genotype groups for participants ≥35yrs old.

	***COL3A1* rs1800255 Genotypes**			**p value [a]**	**p value [c]**
	GG (49)	**GA (53)**	**AA (11)**		
SR ROM (mm)	277.5 ± 126.3 (49)	244.8 ± 91.8 (53)	206.1 ± 104.6 (11)	0.109	0.137
Non-Dom SLR (°)	87.9 ± 24.7 (8)	73.7 ± 19.8 (7)	0.0 ± 0.0 (0)	nd	nd
Dom SLR (°)	92.9 ± 23.9 (8)	81.6 ± 30.7 (7)	0.0 ± 0.0 (0)	nd	nd
Non-Dom ShTR (°)	221.4 ± 21.1 (6)	184.6 ± 28.9 (6)	0.0 ± 0.0 (0)	nd	nd
Dom ShTR (°)	209.1 ± 12.5 (6)	192.3 ± 90.7 (7)	0.0 ± 0.0 (0)	nd	nd
	***COL6A1* rs35796750 Genotypes**			**p value [b]**	**p value [d]**
	TT (44)	**TC (55)**	**CC (13)**		
SR ROM (mm)	253.0 ± 97.7 (44)	260.2 ± 117.2 (55)	277.2 ± 105.9 (13)	0.679	0.258
Non-Dom SLR (°)	92.9 ± 18.6 (8)	76.7 ± 25.7 (9)	0.0 ± 0.0 (0)	nd	nd
Dom SLR (°)	96.6 ± 22.7 (8)	86.6 ± 28.6 (9)	0.0 ± 0.0 (0)	nd	nd
Non-Dom ShTR (°)	203.4 ± 23.7 (7)	199.4 ± 36.5 (7)	0.0 ± 0.0 (0)	nd	nd
Dom ShTR (°)	206.7 ± 23.5 (7)	194.2 ± 24.1 (8)	0.0 ± 0.0 (0)	nd	nd
	***COL12A1* rs970547 Genotypes**			**p value [a]**	**p value [c]**
	AA (57)	**AG (39)**	**GG (4)**		
SR ROM (mm)	261.3 ± 112.1 (57)	246.7 ± 107.7 (39)	302.5 ± 128.4 (4)	0.547	0.658
Non-Dom SLR (°)	95.9 ± 22.3 (9)	72.7 ± 20.9 (6)	64.0 ± 0.0 (1)	nd	nd
Dom SLR (°)	104.7 ± 24.7 (9)	77.3 ± 17.7 (6)	84.0 ± 0.0 (1)	nd	nd
Non-Dom ShTR (°)	215.6 ± 29.2 (8)	192.5 ± 21.5 (4)	183.0 ± 0.0 (1)	nd	nd
Dom ShTR (°)	210.2 ± 26.9 (8)	191.4 ± 15.5 (5)	210.0 ± 0.0 (1)	nd	nd

Values are expressed as mean ± standard deviation. Number of participants is indicated in parentheses. SR, sit-and-reach; mm, millimetres; Non-Dom, non-dominant; SLR, straight leg raise; Dom, dominant; ShTR, shoulder total rotation; nd, not determined.

[a] p values adjusted for gender.

[b] p values adjusted for gender and weight.

[c] p values adjusted for gender and self-reported history of flexibility training.

[d] p values adjusted for gender, weight and self-reported history of flexibility training.

Table 5.7. Age-genotype interaction effects for the range of motion measurements and *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 variants.

Polymorphism	p values				
	SR ROM (mm)	Non-Dom SLR (°)	Dom SLR (°)	Non-Dom ShTR (°)	Dom ShTR (°)
***COL3A1* rs1800255**	0.143	0.643	0.500	0.191	0.741
***COL6A1* rs35796750**	0.476	0.263	0.365	0.147	0.104
***COL12A1* rs970547**	0.573	0.100	0.133	0.279	0.451

SR, sit-and-reach; mm, millimetres; Non-Dom, non-dominant; SLR, straight leg raise; Dom, dominant; ShTR, shoulder total rotation.

5.3.3 Gene-Gene Interactions and Range of Motion

Gene-gene interactions between *COL3A1* rs1800255, *COL6A1* rs35796750, *COL12A1* rs970547, and the previously genotyped *COL5A1* rs12722 [28;29;40], and sit-and-reach ROM (SR-ROM) measurements were investigated. Gene-gene interaction analysis between these variants and the other ROM measures (Non-Dom SLR, Dom SLR, Non-Dom ShTR and Dom ShTR) could not be calculated due to small group sizes. Participants were divided into a High (participants in the upper half of SR-ROM scores) and a Low (participants in the lower half of SR-ROM scores) group based on their SR-ROM scores. No significant gene-gene interactions were

identified when all four gene variants were used to construct pseudo-haplotypes (Figure 5.1), or when any three variants were used (Appendix C.9)

Significant gene-gene interactions were identified when two-gene pseudo-haplotypes were constructed (Appendix C.8). A summary of these significant interactions is presented in table 5.8. The G+C inferred pseudo-haplotype, constructed from *COL3A1* rs1800255 and *COL6A1* rs35796750, was significantly (p=0.017) over-represented in the High SR-ROM group (30.4%, n=52) when compared to the Low SR-ROM group (25.8%, n=44) (Table 5.8). Significant interactions were also identified when *COL5A1* rs12722 and *COL6A1* rs35796750 were analysed. Specifically, the C+T inferred pseudo-haplotype was significantly (p=0.020) over-represented in the High SR-ROM group (27.9%, n=48) when compared to the Low SR-ROM group (23%, n=40), while the T+T inferred pseudo-haplotype was significantly (p=0.028) under-represented in the High SR-ROM group (31.5%, n=54) when compared to the Low SR-ROM group (37.3%, n=64) (Table 5.8). No other significant interactions were identified between any of the variant combinations and SR-ROM groups (Appendix C.8). Gene-gene interactions between these variants and SR-ROM groups in the older group (≥35yrs old) could not be determined due to small sample sizes.

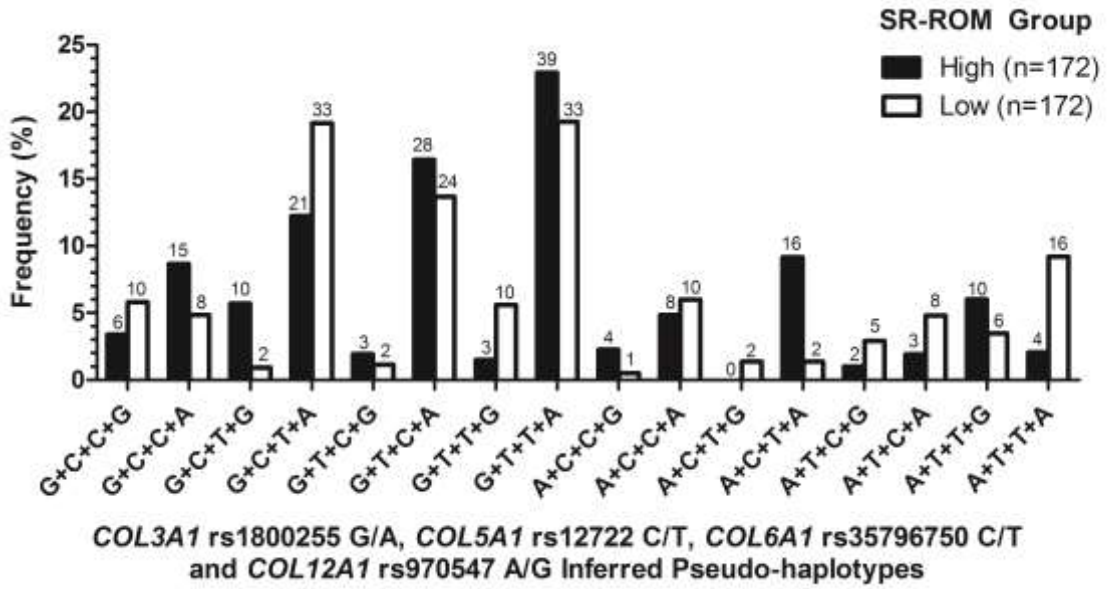


Figure 5.1. Inferred pseudo-haplotype frequencies, constructed from *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs37596750 and *COL12A1* rs970547, and their interaction between High and Low Sit-and-Reach Range of Motion (SR-ROM) Measurements. Numbers of participants are listed above each column.

Table 5.8. A summary of the two-gene inferred pseudo-haplotypes significantly associated with Sit-and-Reach Range of Motion Measurements.

	COL3A1 **rs1800255 (G/A)**	***COL5A1*** **rs12722 (C/T)**	***COL6A1*** **rs35796750 (C/T)**	***COL12A1*** **rs970547 (A/G)**
COL3A1 **rs1800255 (G/A)**			G+C	
COL5A1 **rs12722 (C/T)**			C+T T+T	
COL6A1 **rs35796750 (C/T)**				
COL12A1 **rs970547 (A/G)**				

The gene variant on the left contributes the first allele listed in each combination. Green shading indicates that the inferred pseudo-haplotype is associated with an increased ROM. Red shading indicates that the inferred pseudo-haplotype is associated with a reduced ROM.

5.4 DISCUSSION

The *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 variants are not independently associated with specific upper and lower body joint range of motion (ROM) measurements. Furthermore, no age-genotype effects were identified between these gene variants and any of the measures of ROM investigated in this study. Despite a lack of independent associations, a number of gene-gene interactions were identified. Specifically, two-gene interactions between *COL3A1* rs1800255 and *COL6A1* rs35796750 and *COL5A1* rs12722 and *COL6A1* rs35796750 were associated with sit-and-reach ROM. The G+C inferred pseudo-haplotype, constructed from *COL3A1* rs1800255 and *COL6A1* rs35796750, was associated with increased sit-and-reach ROM and implicates these genes in the aetiology of ROM for the first time. Future studies are therefore required to determine how types III and VI collagen might interact in the collagen fibril and how these interactions would modulate the aetiology of ROM.

Interestingly, the inferred pseudo-haplotypes constructed from *COL5A1* rs12722 and *COL6A1* rs35796750 appeared to be dominated by the rs12722 allele. Specifically within the inferred pseudo-haplotypes, the rs12722 T and C alleles were associated with decreased and increased sit-and-reach ROM respectively, regardless of which *COL6A1* rs35796750 allele was present. Furthermore, these results for *COL5A1* rs12722 are consistent with previously reported findings [28;29]. Interactions between types V and VI have been described [190], however further studies are required to determine the how these interactions might modulate ROM.

Mutations within the four genes investigated in this study result in musculoskeletal soft tissue abnormalities with symptoms of joint hypermobility [17;31;33;109;122-124;186;218]. Future studies are therefore required to investigate additional variants within these genes, especially *COL12A1* which was not implicated in this study, as markers for joint ROM.

As stated above, the results for *COL5A1* rs12722 in this study are consistent with previously reported findings [28;29], however three additional studies have shown contrasting results [13;40;196]. Specifically, the CT genotype was shown to be associated with increased sit-and-reach and straight leg raise ROM in uninjured participants as well as those with Achilles tendon injuries [40], with both increased knee hyperextension and general joint laxity in females only [13], and was over-represented in a cohort of female Italian high-level international rhythmic gymnasts [196]. The average age (±43yrs old) of participants recruited by Collins et al. [40] was considerably higher than that described in this thesis (±32yrs old). Furthermore, approximately 70% (85/119) of the cohort recruited by Collins et al. [40] reported a history of chronic Achilles tendinopathy or Achilles tendon rupture. Therefore, the difference in average ages and injury profiles between these cohorts may explain the inconsistent findings [28;29;40], since it cannot be excluded that the injured participants had an altered ROM at the time of measurement that may not represent their pre-injury values [40]. There are two key differences in methodology between the more recent studies [13;196] and those described in this thesis. (i) Both identified associations between *COL5A1* rs12722 and ROM in younger participants (±22yrs old [13], ±12yrs old [196]), when compared to this study (±32yrs old). (ii) Both identified associations between *COL5A1* rs12722 and ROM in females only [13;196],

whereas this study did not report gender-specific associations. Therefore, one possible explanation for these differences is the change in collagen integrity with increasing age which results in substantial loss of tendon flexibility and reduced ROM [148]. It may therefore be proposed that the difference in type V collagen regulation and accumulation, as a result of rs12722, would become more apparent with increasing age. Future work is required to explore the reasons for these differences. All the studies however agree that variants within the 3'-untranslated region of the *COL5A1* gene are associated with several measures of ROM.

Joint range of motion is an important risk factor for soft tissue injuries [208] and, as will be discussed in the next chapter, is also strongly associated with running economy [77]. Therefore, it is important to identify the intrinsic and extrinsic factors that may determine range of motion, including the genetic component. This would allow for better understanding of the aetiologies of both injury risk and performance. As such additional research is required to further identify factors associated with joint range of motion. This will be discussed in more detail in the final concluding chapter of this thesis (Chapter 10).

As expected, female participants had higher measurements for all measures of range of motion than male participants [6;73]. Participant age did not correlate with range of motion measurements. This is in contradiction to previously published findings [44;90]. A possible explanation for this is that 150 participants (43%) were below the age of 25 yrs. Furthermore, in agreement with previous studies [46;193], sit-and-reach range of motion measurements were significantly different between participants with a self-reported history of flexibility training when compared to those

with no history of flexibility training, and both dominant and non-dominant leg straight leg raise range of motion measurements were also higher in participants with self-reported flexibility training. No significant differences in dominant or non-dominant total shoulder rotation range of motion measurements were identified between participants with a self-reported history of flexibility training when compared to those with no history of flexibility training. This may be due to the low number (12.3%, n=43) of participants that reported doing upper body flexibility training. In addition, the type of sporting or other physical activities performed by these participants may also be responsible for a lack of difference between dominant or non-dominant total shoulder rotation range of motion measurements.

One limitation to this study was that the state of menstrual cycle at the time of recruitment and ROM testing was not recorded for the female participants. Furthermore, the number of peri- and/or postmenopausal women was also not recorded. The effects of menstrual cycle on ROM, measured by knee joint laxity, have previously produced equivocal results [214].

In conclusion, a novel gene-gene interaction between *COL3A1* rs1800255 and *COL6A1* rs35796750 was significantly associated with sit-and-reach ROM. In addition, significant interactions were also identified between *COL5A1* rs12722 and *COL6A1* rs35796750. No independent associations or age-genotype effects were determined between *COL3A1* rs1800255, *COL6A1* rs35796750 or *COL12A1* rs970547 and measures of upper and lower body range of motion. Additional variants within these genes, as well as additional genes, should be investigated for their contribution to inter-individual differences in joint range of motion. This would

add to the current understanding of range of motion as well as other phenotypes which are known to be heavily modulated by changes in range of motion, such as soft tissue injury [208] and athletic performance [77]. A summary of the results of this chapter are presented in table 5.9.

Table 5.9. A continuing summary of the results of the chapters in this thesis. The results of chapter 5, and all preceding chapters, are shown.

Investigated Phenotype	Associations	Investigated Gene Variants			
		COL3A1 rs1800255 (G/A)	***COL5A1*** rs12722 (C/T)	***COL6A1*** rs35796750 (C/T)	***COL12A1*** rs970547 (A/G)
ACL	**Genotype**	not associated	CC ♀ prev pub[1]	not associated	AA ♀ prev pub[2]
	Haplotype	nd	T ♀[3]	nd	A ♀[3]
Achilles tendinopathy	**Genotype**	not associated prev pub[4]	CC prev pub[5]	not associated	not associated prev pub[6]
	Haplotype	A	C	C	nd
ROM	**Genotype**	not associated	CC prev pub[7]	not associated	not associated
	Haplotype	not associated	C	T	not associated
Endurance Performance		Chapter 6			
EAMC		Chapter 7			
Rugby Union		Chapter 8			

Green shading indicates that the allele or genotype is associated with a reduced risk of injury, except for ROM where it represents an increased ROM when compared to other genotypes. Red shading indicates that the allele or genotype is associated with an increased risk of injury. Grey shading indicates no independent association. Prev pub, previously published. ♀, associations identified only in females. ACL, anterior cruciate ligament; ROM, range of motion; EAMC, exercise-associated muscle cramps; nd, not determined.

[1] independent association previously reported by Posthumus et al. [153].

[2] independent association previously reported by Posthumus et al. [154], but not independently associated in this thesis.

[3] T+A pseudo-haplotype also associated with increased risk of female ACL ruptures when a Polish cohort was analysed [137].

[4] no independent association previously reported by Saunders [170]

[5] independent association previously reported by Mokone et al. [131] and repeated by September et al. [179].

[6] no independent association previously reported by September et al. [181].

[7] independent association previously reported by Brown et al. [28;29].

CHAPTER 6

GENETIC MARKERS FOR ENDURANCE PERFORMANCE

The data presented in this chapter has been published in a condensed form in the peer-reviewed articles:
O'Connell K, Posthumus M, Collins M. *COL6A1* gene and Ironman triathlon performance. *International Journal of Sports Medicine.* 2011;32(11):896-901.
O'Connell K, Posthumus M and Collins M. Collagen gene interactions and endurance running performance. *South African Journal of Sports Medicine*. 2014;26(1):9-14

6.1 INTRODUCTION

It is widely accepted that multiple extrinsic factors, such as training and nutrition, play a role in determining athletic ability [118]. However, studies have also shown that athletic ability is also modulated by numerous intrinsic factors. Inter-individual genetic variation [26] and joint range of motion (ROM) [45;77] are two intrinsic factors associated with athletic ability. Reduced ROM, specifically sit-and-reach ROM, has also been associated with improved endurance running economy and performance [45;77;82]. As described previously in chapter 2 section 2.3.2 and in chapter 5 section 5.1, the *COL5A1* rs12722 variant was associated with lower limb ROM measurements, including sit-and-reach ROM in numerous studies [29;40;196]. Furthermore, the *COL5A1* rs12722 TT genotype was associated with improved

endurance running performance during the South African Ironman triathlon [151]. The association between the *COL5A1* rs12722 TT and rs71746744 (-/AGGG) AGGG/AGGG genotypes and improved endurance running performance was later replicated in a road running event [1;28]. In addition, as described in chapter 2 section 2.2, these functional *COL5A1* variants regulate type V collagen production [107], which may lead to increased type V collagen production [107], thereby affecting normal collagen fibrillogenesis and altering the mechanical properties of the tissue, leading to improved endurance performance [41].

Similarly to types V collagen, types III, VI and XII collagen also regulate fibrillogenesis [56;115;129;141;213], and by implication the mechanical properties of connective tissue, resulting in inter-individual variations in endurance performance. Therefore, like the *COL5A1* rs12722 gene variant, it may be proposed that common potentially functional variants within the *COL3A1*, *COL6A1* and *COL12A1* genes, such as the *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 variants investigated in the previous chapters of this thesis, may also be associated with athletic endurance performance.

Furthermore, as mentioned in chapter 2 section 2.6.4, mutations within the *COL6A1* gene cause severe skeletal muscle disorders such as Bethlem myopathy [17;109] and Ullrich congenital muscular dystrophy [17;109;140], characterized by muscle weakness and connective tissue abnormalities. A murine *Col6a1* knockout model, that simulates these myopathies, revealed that the average weekly distance run by these myopathic *Col6a1 -/-* mice was consistently less than the wildtype mice [21],

strengthening the hypothesis that common variants within *COL6A1* are associated with athletic performance.

Since *COL5A1* rs12722 was previously associated with endurance running performance, the primary aim of this study was to determine if *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 are associated with athletic endurance performance in participants of four South African Ironman triathlon events. Specifically, it may be hypothesised that, due to the proposed functional effects of these variants, the *COL6A1* rs35796750 TT and *COL12A1* rs970547 AA genotypes, and either *COL3A1* rs1800255 genotype, would be associated with improved endurance performance. The secondary aim of this study was to investigate gene-gene interactions between *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547, and *COL5A1* rs12722 where appropriate, and endurance performance. Specifically, that the *COL5A1* rs12722 T, *COL6A1* rs35796750 T, *COL12A1* rs970547 A and either of the *COL3A1* rs1800255 alleles may be implicated in gene-gene interactions associated with increased athletic endurance performance.

6.2 MATERIALS AND METHODS

6.2.1 Participants

A total of six hundred and sixty one male Caucasian participants were recruited from four South African Ironman triathlon events, consisting of a 3.8 km sea swim, 180 km cycle and 42.2 km run, for this genetic case-control association study. Specifically, participants were recruited at the registration of either the 2000 (n=96) and 2001 (n=294) events held in Gordon's Bay (approximately 50km outside of Cape Town) or the 2006 (n=219) and 2007 (n=52) Port Elizabeth events (approximately 750km east of Cape Town). Only male participants were recruited since gender is a known factor in determining athletic performance [26;198].

Race results were obtained from the race organisers and participants were divided into three equal tertiles, for the 3.8 km swim, the 180 km bike and the 42.2 km run components, based on their finishing times. The fastest triathletes were placed into the Fast tertile. Those that finished in the mid-field were placed in the Middle tertile and the slowest triathletes were placed into the Slow tertile.

All participants completed informed consent forms (Appendix A.3) and a physical activity questionnaire (Appendix A.4). In addition, participants of the Port Elizabeth sub-group completed training history questionnaires (Appendix A.4). Training history was not documented at the events held in Gordon's Bay. Approval for the study was obtained from the Human Research Ethics Committee of the Faculty of Health

Sciences within the University of Cape Town and the race organisers of each event (Appendix A.1).

For participants that entered more than one event only data from one race year was used. Since training data was obtained during the Port Elizabeth events, the event priority for the participants who had completed more than one event was 2006, then 2007 and finally 2001, which had a more complete and larger dataset than the 2000 event.

6.2.2 Variant Selection

Three independent sequence variants were investigated in this study. Namely *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547. The reasons for their selection have been stated in chapter 3, section 3.2.2. Furthermore, these variants have been independently associated with or interact to (i) modulate the risk of musculoskeletal soft tissue injuries (Chapter 3 and 4) and/or (ii) modulate joint range of motion (Chapter 5).

6.2.3 DNA Extraction and Genotyping Methods

Approximately 4.5ml of venous blood was collected, at the registration of each event, from all participants by venipuncture of a forearm vein into an EDTA vacutainer tube. Blood samples were stored at 4˚C until DNA was extracted, as previously described [108], with minor modifications [131]. An outline of this methodology was described

in chapter 3, section 3.2.3. All analyses were performed at the UCT/MRC Research Unit for Exercise Science and Sports Medicine, University of Cape Town, South Africa.

Genotyping of *COL3A1* rs1800255 and *COL6A1* rs35796750 was performed using custom designed Fluorescence-based Taqman® polymerase chain reaction (PCR) assays (Applied Biosystems, Foster City, CA, USA) as described in chapter 3. *COL12A1* rs970547 was genotyped using the *Alu*I restrinction fragment length polymorphism (RFLP) method as previously described in chapter 5 [154;181]. Additional genotyping information is available in appendix B. All genotyping was done at the UCT/MRC Research Unit for Exercise Science and Sports Medicine, University of Cape Town, Cape Town, South Africa.

Investigators were blinded to the performance results of the participant samples while genotyping. Two investigators independently confirmed genotyping. Furthermore, a number of positive and negative controls were used to ensure genotyping accuracy. No discrepancies were observed. A total of 642/661 (97.1%), 661/661 (100.0%) and 629/661 (95.2%) participants were genotyped for *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547.

6.2.4 Statistical Analysis

The same statistical tests and haplotype analysis methods were performed as described in chapter 3 section 3.2.4. In addition the magnitude of changes in performance variables were determined on a scale of effect sizes (ES) where: <0.2 = trivial, 0.21-0.6 = small, 0.61-1.2 = moderate, 1.21-2.0 = large, 2.1-4.0 = very large and >4.0 = nearly perfect [76]. Bivariate correlations were used to determine the relationship between overall race time and participant physiological characteristics. Multivariate analyses were performed by forward stepwise regression. Significance was accepted when $p<0.05$. No adjustments were made for multiple testing in this study for the same reasons as outlined in chapter 3 section 3.2.4.

6.3 RESULTS

6.3.1 Participant Characteristics

The general characteristics of the six hundred and sixty one male participants included in this study are presented in table 6.1. Participants from the Gordon's Bay events (2000 and 2001; 34.7 ± 7.9 years, n=390) were significantly younger ($p<0.001$) than those recruited from the Port Elizabeth events (2006 and 2007; 38.2 ± 8.4 years, n=269) (Table 6.1). Significantly less ($p=0.002$ and $p=0.037$) South African born participants competed in the 2000 (53.2%, n=50) and 2001 (63.6%, n=185) events when compared to the 2006 event (73.3%, n=118) (Table 6.1). Significantly less ($p=0.002$) participants residing in South Africa competed in the

2001 event (75.7%, n=215) when compared to the 2006 event (86.8%, n=190) (Table 6.1).

Participant age (r=0.220, p<0.001), weight (r=0.318, p<0.001) and BMI (r=0.394, p<0.001) were all significantly correlated with the overall race time (Figure 6.1). Participants height was not significantly correlated with overall race time (r=0.007, p=0.865) (Figure 6.1C) In addition mean overall race time was significantly different between the race years (p<0.001) (Table 6.1).

No genotype effects were identified between any of the participant characteristics and the *COL3A1* rs1800255, *COL6A1* rs35796750 or *COL12A1* rs970547 variants (Table 6.2). The *COL3A1* rs1800255 (p=0.428), *COL6A1* rs35796750 (p=0.178) and *COL12A1* rs970547 (p=0.062) variants were in Hardy-Weinberg Equilibrium.

Table 6.1. General characteristics for the South African Ironman triathlon participants that were recruited at the registration of either the 2000 and 2001 events held in Gordon's Bay or the 2006 and 2007 Port Elizabeth events.

	All (661)	**2000 event (96)**	**2001 event (294)**	**2006 event (219)**	**2007 event (52)**	**p-value**
Age (yrs)	36.1 ± 8.3 (659)	34.5 ± 7.2 (96)	34.7 ± 8.1 (294)	38.2 ± 8.6 (219)	38.4 ± 7.1 (50)	**<0.001**
Height (cm)	180.5 ± 6.6 (559)	180.5 ± 7.4 (85)	180.5 ± 6.5 (267)	180.3 ± 6.4 (158)	181.4 ± 6.8 (49)	0.794
Weight (kg)	78.6 ± 9.4 (586)	77.5 ± 10.2 (94)	78.8 ± 8.7 (274)	78.2 ± 9.3 (166)	80.9 ± 11.2 (52)	0.196
BMI ($kg.m^{-2}$)	24.0 ± 2.3 (555)	23.7 ± 2.4 (85)	24.0 ± 2.1 (264)	24.0 ± 2.2 (157)	24.7 ± 3.2 (49)	0.084
CoB (% SA)	65.0 (388)	53.2 (50)	63.6 (185)	73.3 (118)	68.6 (35)	**0.011**
CoR (%SA)	81.2 (468)	85.7 (12)	75.7 (215)	86.8 (190)	86.3 (44)	**0.011**
Overall (min)	768.6 ± 96.0 (661)	750.3 ± 94.3 (96)	755.9 ± 93.9 (294)	787.6 ± 94.8 (219)	794.5 ± 100.0 (52)	**<0.001**

Values are expressed as mean ± standard deviations or as percentages where appropriate. Number of participants is indicated in parentheses. Values in bold typeset are significant ($p<0.05$). BMI, body mass index. CoB, country of birth. CoR, country of residence. SA, South Africa. Overall, overall race time.

Table 6.2. Genotype effects of *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 on physiological characteristics of participants.

Polymorphism	p-values					
	Age	**Height**	**Weight**	**BMI**	**CoB**	**CoR**
***COL3A1* rs1800255**	0.852	0.293	0.347	0.402	0.630	0.859
***COL6A1* rs35796750**	0.341	0.627	0.948	0.332	0.164	0.681
***COL12A1* rs970547**	0.463	0.319	0.903	0.438	0.866	0.829

BMI, body mass index. CoB, country of birth. CoR, country of residence

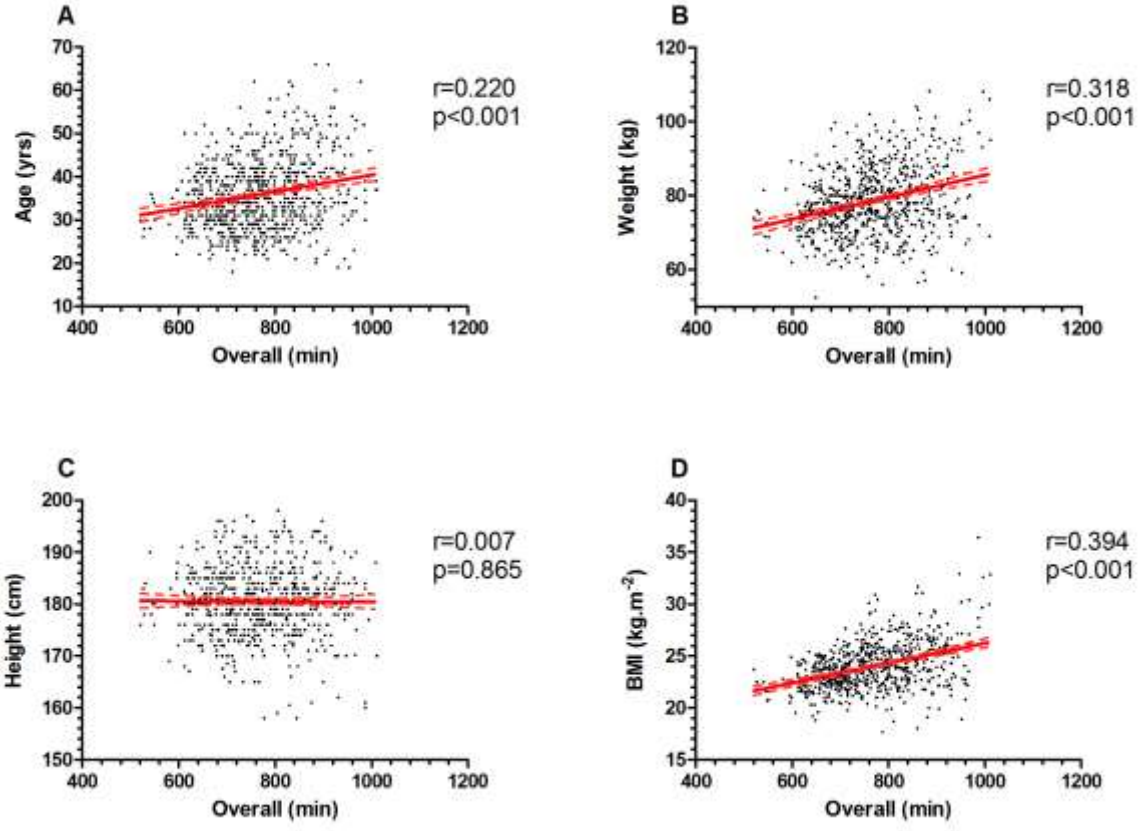


Figure 6.1. Correlations between overall race time and participant (A) age, (B) weight, (C) height and (D) body mass index (BMI).

6.3.2 Participant Training History

Participant self-reported training history data, characterising the 15 weeks prior to each event, collected at the 2006 and 2007 Port Elizabeth South African Ironman triathlon events is presented in table 5.3. The distance (km/wk) and duration (hrs/wk) trained, for the swim, bike and run, and combined, were similar for each *COL6A1* rs35796750 genotype group. Interestingly, although not significant, linear trends were observed in the self-reported bike training duration (p=0.059) and run training duration (p=0.050) between genotype groups (Table 6.3). Specifically, for bike training duration, triathletes with the *COL6A1* rs35796750 TT genotype trained for the least number of hours per week (7.7 ± 2.6 hrs/wk) followed by those with a TC genotype (8.1 ± 2.8 hrs/wk), while triathletes with a CC genotype (9.1 ± 3.6 hrs/wk) trained the most hours per week. The magnitude of these changes in genotype specific duration cycled were considered “small” (ES=0.24). Similarly, for run training duration, triathletes with the *COL6A1* rs35796750 TT genotype trained for the least number of hours per week (4.5 ± 1.5 hrs/wk) followed by those with a TC genotype (4.6 ± 1.7 hrs/wk), while triathletes with a CC genotype (5.9 ± 5.9 hrs/wk) trained the most hours per week. The magnitude of these changes in genotype specific duration run were also considered “small” (ES=0.16).

Although probably not biologically relevant, the *COL3A1* rs1800255 variant was significantly (p=0.002) associated with swim training duration (hrs/wk). Specifically, participants with a *COL3A1* rs1800255 GA genotype (3.4 ± 1.6 hrs/wk) trained significantly (p=0.001) more hours than those participants with a *COL3A1* rs1800255 GG (2.8 ± 1.0 hrs/wk) or AA (2.3 ± 0.9 hrs/wk) genotype. The magnitude of these changes in genotype specific swimming duration were considered “small” (ES=0.48).

The distance (km/wk) and duration (hrs/wk) trained, for the bike, run and combined, were not significantly associated with *COL3A1* rs1800255 (Table 6.3). Furthermore, no significant associations were identified between *COL12A1* rs970547 and distance or duration trained, for the swim, bike, run or combined (Table 6.3).

Table 6.3. Participant self-reported training history for the *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 genotype groups of the Port Elizabeth sub-group.

	All (187)	***COL3A1* rs1800255 Genotype**			p value
		GG (97)	**GA (73)**	**AA (17)**	
Training (km/wk)					
Swim	6.4 ± 3.0 (185)	6.4 ± 2.8 (95)	6.5 ± 3.2 (73)	6.1 ± 3.1 (17)	0.857
Bike	224.3 ± 84.9 (170)	218.2 ± 92.9 (85)	232.0 ± 76.5 (69)	223.8 ± 76.2 (16)	0.606
Run	45.7 ± 18.0 (182)	47.0 ± 20.4 (92)	44.8 ± 13.3 (73)	42.4 ± 21.6 (17)	0.535
Combined [a]	236.9 ± 85.8 (155)	230.2 ± 95.6 (76)	245.3 ± 75.2 (63)	235.3 ± 77.3 (16)	0.588
Training (hrs/wk)					
Swim	3.0 ± 1.3 (184)	2.8 ± 1.0 (97)	3.4 ± 1.6 (70)	2.3 ± 0.9 (17)	**0.002**
Bike	8.1 ± 2.9 (171)	8.1 ± 3.2 (89)	8.2 ± 2.5 (66)	7.8 ± 2.9 (16)	0.875
Run	4.5 ± 1.7 (174)	4.6 ± 1.9 (92)	4.5 ± 1.4 (65)	4.1 ± 1.8 (17)	0.575
Combined [a]	15.4 ± 4.8 (162)	15.5 ± 4.8 (85)	15.6 ± 3.7 (61)	14.3 ± 4.5 (16)	0.568
	All (271)	***COL6A1* rs35796750 Genotype**			**p value**
		TT (76)	**TC (139)**	**CC (56)**	
Training (km/wk)					
Swim	6.4 ± 3.0 (215)	6.6 ± 3.0 (58)	6.4 ± 3.1 (110)	6.3 ± 2.8 (47)	0.820
Bike	224.7 ± 82.6 (199)	224.0 ± 78.6 (54)	221.1 ± 81.9 (102)	234.1 ± 89.9 (43)	0.689
Run	47.0 ± 18.1 (212)	46.4 ± 14.0 (58)	45.5 ± 17.1 (108)	51.5 ± 23.9 (46)	0.162
Combined [a]	238.0 ± 82.5 (183)	233.8 ± 77.0 (49)	237.5 ± 83.3 (93)	244.1 ± 88.5 (41)	0.840
Training (hrs/wk)					
Swim	3.1 ± 1.7 (214)	3.2 ± 1.8 (59)	2.9 ± 1.2 (112)	3.3 ± 3.4 (43)	0.340
Bike	9.0 ± 12.5 (198)	7.7 ± 2.6 (55)	8.1 ± 2.8 (103)	13.2 ± 27.2 (40)	0.059
Run	4.9 ± 3.1 (201)	4.5 ± 1.5 (54)	4.6 ± 1.7 (106)	5.9 ± 5.9 (41)	0.050
Combined [a]	15.6 ± 4.4 (188)	14.9 ± 3.9 (53)	15.5 ± 4.3 (99)	16.8 ± 5.2 (36)	0.148
	All (207)	***COL12A1* rs970547 Genotype**			**p value**
		AA (120)	**AG (82)**	**GG (5)**	
Training (km/wk)					
Swim	6.4 ± 3.0 (207)	6.3 ± 2.8 (120)	6.7 ± 3.2 (82)	5.6 ± 1.5 (5)	0.531
Bike	222.7 ± 83.3 (190)	216.9 ± 83.6 (110)	231.6 ± 82.8 (75)	218.7 ± 90.4 (5)	0.499
Run	46.3 ± 17.8 (203)	44.9 ± 16.6 (117)	48.5 ± 19.6 (81)	44.0 ± 9.6 (5)	0.371
Combined [a]	236.5 ± 83.6 (174)	229.8 ± 82.6 (97)	246.1 ± 84.8 (72)	229.5 ± 89.9 (5)	0.449
Training (hrs/wk)					
Swim	3.0 ± 1.5 (206)	3.0 ± 1.7 (119)	3.1 ± 1.4 (82)	2.3 ± 0.7 (5)	0.513
Bike	8.9 ± 12.8 (190)	9.5 ± 16.6 (110)	8.2 ± 2.9 (75)	7.6 ± 2.1 (5)	0.752
Run	4.9 ± 3.1 (194)	4.9 ± 3.7 (113)	4.8 ± 2.0 (76)	4.3 ± 0.4 (5)	0.896
Combined [a]	15.5 ± 4.4 (180)	15.3 ± 4.3 (103)	15.8 ± 4.7 (72)	14.2 ± 2.3 (5)	0.607

Values are expressed as mean ± standard deviations. Number of participants is indicated in parentheses. Values in bold typeset are significant ($p<0.05$).

[a] Combined = Swim + Bike + Run

6.3.4 Collagen Genes and Performance in the South African Ironman triathlon

The finishing times for the swim, bike and run components of the Ironman triathlon were independently compared between the *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 genotype groups. Overall finishing time was not analysed since any identified associations would merely be due to component specific associations. Significant *COL6A1* rs35796750 genotype differences were observed in the times to complete the bike (p=0.033) component when all participants were analysed (Table 6.4). Specifically, participants with the *COL6A1* rs35796750 TT genotype completed the bike component significantly (p=0.014) faster than those with a CC or TC genotype. The magnitude of the change in bike performance between triathletes with the *COL6A1* rs35796750 TT genotype compared to those with a TC or CC genotype was considered “small” (ES=0.21). No significant genotype effects were identified for the swim or run finishing times. The *COL3A1* rs1800255 and *COL12A1* rs970547 variants were not significantly associated with time to complete any of the individual components (3.8 km swim, 180 km bike or 42.2 km run) of the 226 km Ironman triathlon (Table 6.4).

Table 6.4. Finishing times for the *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 genotype groups in the 3.8km swim, 180km bike and 42.2km run components.

	All (642)	***COL3A1* rs1800255 Genotype**			p value
		GG (333)	**GA (265)**	**AA (44)**	
3.8 km Swim (min)	77.2 ± 17.4 (629)	77.4 ± 17.3 (326)	76.6 ± 17.6 (259)	79.7 ± 17.2 (44)	0.535
180 km Bike (min)	393.7 ± 42.0 (615)	394.0 ± 42.7 (317)	392.2 ± 41.6 (254)	399.4 ± 39.9 (44)	0.565
42.2 km Run (min)	288.4 ± 49.1 (620)	285.9 ± 50.2 (324)	289.9 ± 47.4 (253)	297.8 ± 51.1 (43)	0.264
	All (661)	***COL6A1* rs35796750 Genotype**			**p value**
		TT (196)	**TC (344)**	**CC (121)**	
3.8 km Swim (min)	77.9 ± 17.4 (644)	76.0 ± 16.4 (192)	78.5 ± 17.7 (337)	79.1 ± 18.0 (115)	0.191
180 km Bike (min)	394.6 ± 42.1 (630)	388.3 ± 40.6 (186)	398.3 ± 42.4 (331)	394.4 ± 42.8 (113)	**0.033**
42.2 km Run (min)	288.2 ± 50.1 (640)	284.9 ± 51.1 (193)	291.0 ± 49.7 (331)	285.8 ± 49.4 (116)	0.348
	All (629)	***COL12A1* rs970547 Genotype**			**p value**
		AA (344)	**AG (255)**	**GG (30)**	
3.8 km Swim (min)	78.5 ± 17.6 (614)	78.8 ± 17.1 (334)	78.2 ± 17.9 (251)	76.9 ± 20.4 (29)	0.800
180 km Bike (min)	395.2 ± 41.5 (600)	393.8 ± 39.8 (330)	397.6 ± 43.2 (243)	391.6 ± 47.3 (27)	0.504
42.2 km Run (min)	290.1 ± 49.7 (609)	289.0 ± 50.7 (331)	290.8 ± 49.3 (249)	297.0 ± 40.7 (29)	0.681

Values are expressed as mean ± standard deviations. Number of participants is indicated in parentheses. Values in bold typeset are significant ($p<0.05$).

Participants were grouped into Fast, Middle and Slow performance tertiles of equal number. Significant (p=0.009) differences in *COL6A1* rs35796750 genotype frequency distribution were identified between the bike component tertiles (Table 6.6). When the *COL6A1* rs35796750 TT genotype was compared to the C allele (TC+CC genotypes), a significant linear trend (p=0.008) was identified in the bike performance tertiles. Specifically, the *COL6A1* rs35796750 TT genotype had the highest representation in the Fast tertile (TT genotype 35.7%), followed by the Middle tertile (TT genotype 29.0%) and the Slow tertile (TT genotype 23.8%) (Figure 6.2). No significant linear trends were identified between *COL6A1* rs35796750 and the swim (p=0.463) or run (p=0.741) performance tertiles (Table 6.6).

No significant differences were identified for *COL3A1* rs1800255 (Table 6.5) or *COL12A1* rs970547 (Table 6.7) between the tertiles for time to complete any of the individual components of the Ironman triathlon.

Table 6.5. Performance tertiles for the *COL3A1* rs1800255 genotype groups in the 3.8km swim, 180km bike and 42.2km run components of the triathlon.

	***COL3A1* rs1800255 Genotype**			**p value**
3.8 km Swim	**GG (333)**	**GA (265)**	**AA (44)**	
Fast	52.7 (119)	40.7 (92)	6.6 (15)	
Middle	50.7 (104)	43.4 (89)	5.9 (12)	0.803
Slow	52.0 (103)	39.4 (78)	8.6 (17)	
	***COL3A1* rs1800255 Genotype**			
180 km Bike	**GG (333)**	**GA (265)**	**AA (44)**	
Fast	53.0 (115)	41.9 (91)	5.1 (11)	
Middle	48.4 (103)	42.7 (91)	8.9 (19)	0.506
Slow	53.5 (99)	38.9 (72)	7.6 (14)	
	***COL3A1* rs1800255 Genotype**			
42.2 km Run	**GG (333)**	**GA (265)**	**AA (44)**	
Fast	56.4 (119)	38.4 (104)	5.2 (11)	
Middle	50.0 (104)	41.8 (87)	8.2 (17)	0.565
Slow	50.3 (101)	42.3 (85)	7.5 (15)	

Values are expressed as percentages. Number of participants is indicated in parentheses.

Table 6.6. Performance tertiles for the *COL6A1* rs35796750 genotype groups in the 3.8km swim, 180km bike and 42.2km run components of the triathlon.

	***COL6A1* rs35796750 Genotype**			**p value**
3.8 km Swim	**TT (196)**	**TC (344)**	**CC (121)**	
Fast	32.0 (73)	51.8 (118)	16.2 (37)	
Middle	30.0 (63)	49.1 (103)	21.0 (44)	0.463
Slow	27.2 (56)	56.3 (116)	16.5 (34)	
	***COL6A1* rs35796750 Genotype**			
180 km Bike	**TT (196)**	**TC (344)**	**CC (121)**	
Fast	35.7 (75)	42.9 (90)	21.4 (450	
Middle	28.6 (60)	57.1 (120)	14.3 (30)	**0.009**
Slow	24.3 (51)	58.1 (122)	17.6 (37)	
	***COL6A1* rs35796750 Genotype**			
42.2 km Run	**TT (196)**	**TC (344)**	**CC (121)**	
Fast	29.7 (66)	50.0 (111)	20.3 (45)	
Middle	31.3 (67)	50.5 (108)	18.2 (39)	0.741
Slow	29.4 (60)	54.9 (112)	15.7 (32)	

Values are expressed as percentages. Number of participants is indicated in parentheses. Values in bold typeset are significant ($p<0.05$).

Table 6.7. Performance tertiles for the *COL12A1* rs970547 genotype groups in the 3.8km swim, 180km bike and 42.2km run components of the triathlon.

	***COL12A1* rs970547 Genotype**			**p value**
3.8 km Swim	**AA (344)**	**AG (255)**	**GG (30)**	
Fast	54.8 (115)	41.0 (86)	4.2 (9)	
Middle	50.8 (102)	44.2 (89)	5.0 (10)	0.704
Slow	57.6 (117)	37.4 (76)	5.0 (10)	
	***COL12A1* rs970547 Genotype**			
180 km Bike	**AA (344)**	**AG (255)**	**GG (30)**	
Fast	55.9 (114)	39.2 (80)	4.9 (10)	
Middle	56.7 (118)	39.9 (83)	3.4 (7)	0.796
Slow	52.4 (98)	42.3 (80)	5.3 (10)	
	***COL12A1* rs970547 Genotype**			
42.2 km Run	**AA (344)**	**AG (255)**	**GG (30)**	
Fast	56.4 (114)	40.1 (81)	3.5 (7)	
Middle	56.5 (117)	39.1 (81)	4.4 (9)	0.461
Slow	50.0 (100)	43.5 (87)	6.5 (13)	

Values are expressed as percentages. Number of participants is indicated in parentheses.

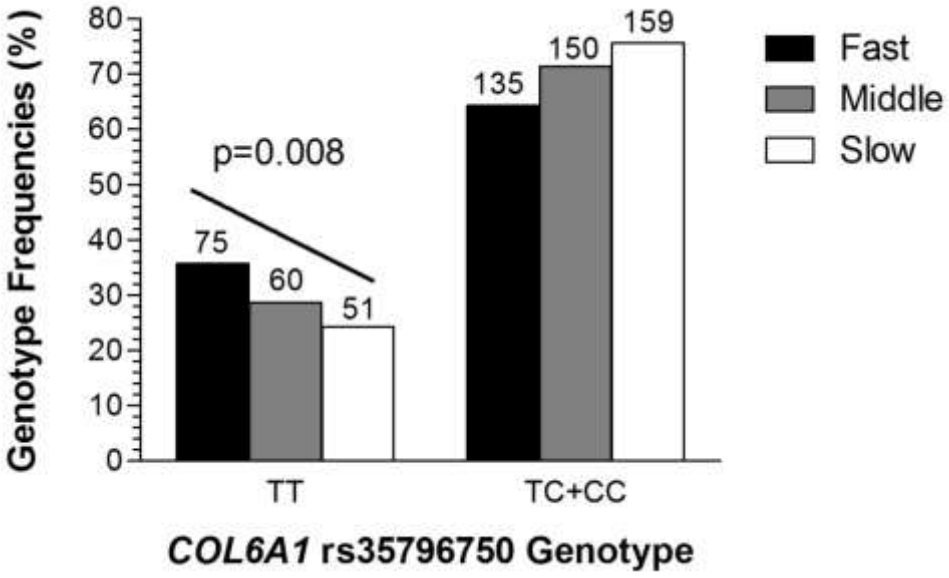


Figure 6.2. Genotype frequency distributions (TT vs. TC+CC) for the *COL6A1* rs35796750 T/C variant in the Fast, Middle and Slow tertiles of the bike component of the triathlon. Number of participants (n) is indicated above each specific column.

6.3.4 Gene-Gene Interactions and Performance in the South African Ironman triathlon

Inferred pseudo-haplotypes between *COL3A1* rs1800255 G/A, *COL6A1* rs35796750 C/T and *COL12A1* rs970547 A/G were constructed for the swim, cycle and run times (Figure 6.3). When the swim component of the triathlon was analysed, the A+C+A inferred pseudo-haplotype was significantly (p=0.047) under-represented in the Fast swim tertile (5.4%, n=14) when compared to the Slow swim tertile (12.6%, n=29) (Figure 6.3A). No significant gene-gene interactions were identified when the cycling component of the triathlon was identified (Figure 6.3B). Analysis of the run component showed that the G+C+A inferred pseudo-haplotype was significantly (p=0.049) over-represented in the Fast run tertile (26.1%, n=65) when compared to the Slow run tertile (20.1%, n=46) (Figure 6.3C).

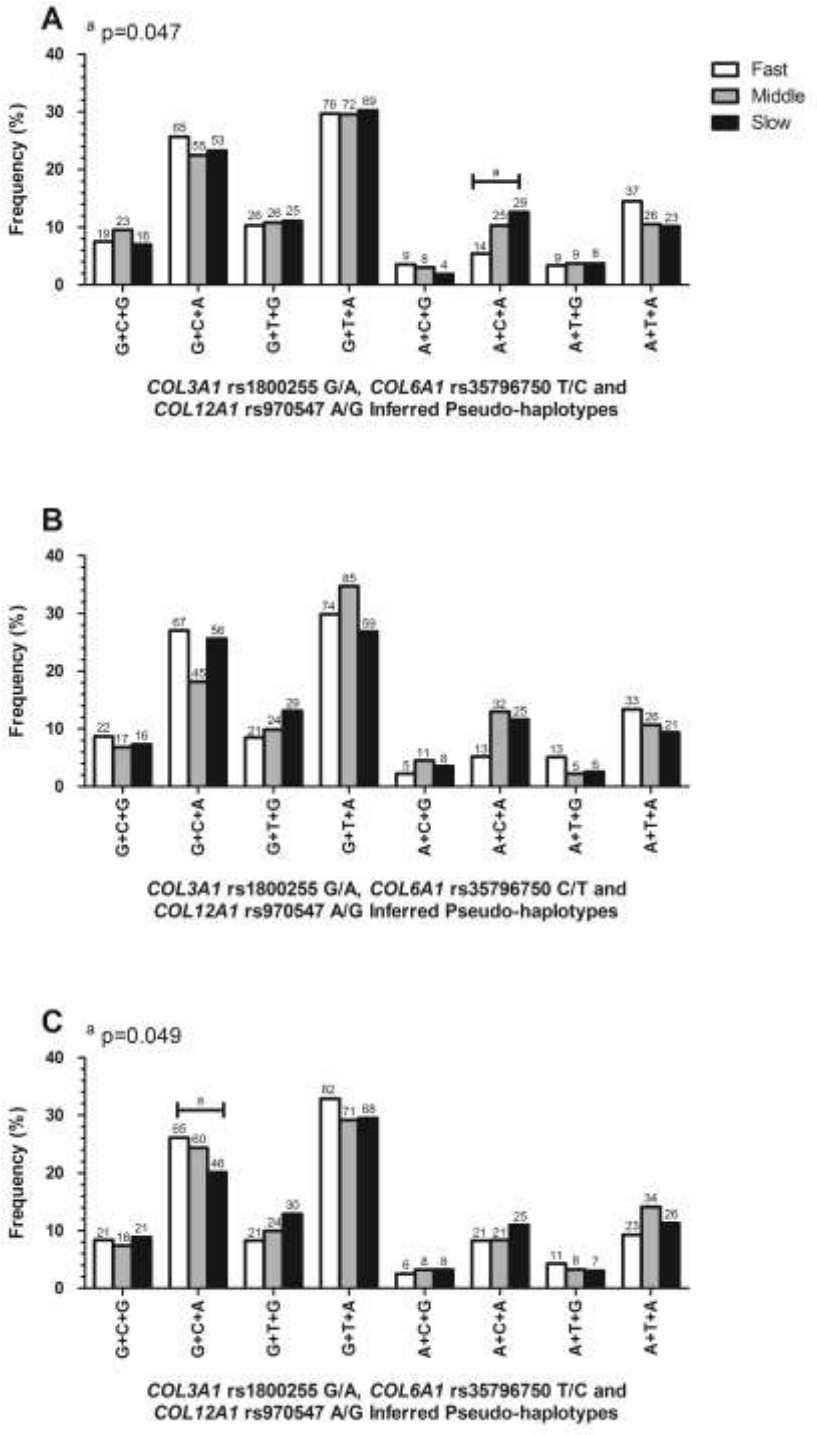


Figure 6.3. Frequency distributions of inferred pseudo-haplotypes constructed from *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 between the Fast, Middle and Slow tertiles in the time to complete (A) the swim component of the triathlon, (B) the cycling component of the triathlon and (C) the run component of the triathlon. Number of participants is indicated above each column.

Since *COL5A1* rs12722 was previously associated with endurance running performance in this cohort [151], inferred pseudo-haplotypes were constructed from *COL3A1* rs1800255 G/A, *COL5A1* rs12722 C/T, *COL6A1* rs35796750 C/T and *COL12A1* rs970547 A/G for the three disciplines of the Ironman triathlon (Figure 6.4). When the swim component of the triathlon was analysed, the G+C+T+A inferred pseudo-haplotype was significantly (p=0.020) over-represented in the Fast swim tertile (15.1%, n=39) when compared to the Slow swim tertile (8.2%, n=19), while the A+T+C+A inferred pseudo-haplotype was significantly (p=0.028) under-represented in the Fast swim tertile (2.5%, n=6) when compared to the Slow swim tertile (8.8%, n=20) (Figure 6.4A). No significant gene-gene interactions were identified when the cycling component of the triathlon was identified (Figure 6.4B). Analysis of the run component showed that the G+T+C+A inferred pseudo-haplotype was significantly (p=0.043 and p=0.025) over-represented in both the Fast (15.7%, n=39) and Middle (17.6%, n=43) run tertiles when compared to the Slow run tertile (11.5%, n=27) respectively (Figure 6.4C).

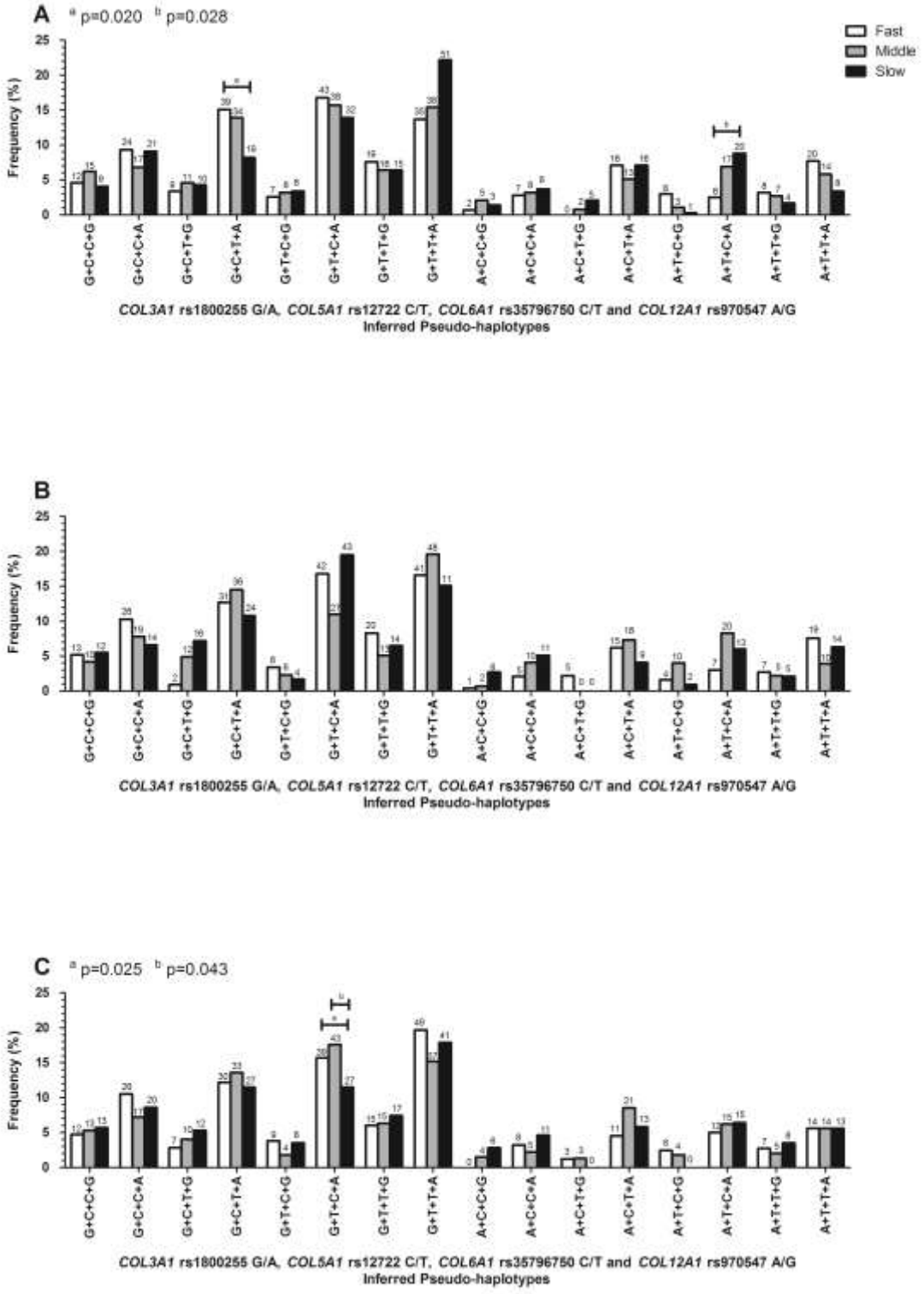


Figure 6.4. Frequency distributions of inferred pseudo-haplotypes constructed from *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 between the Fast, Middle and Slow tertiles in the time to complete (A) the swim component of the triathlon, (B) the cycling component of the triathlon and (C) the run component of the triathlon. Number of participants is indicated above each column.

Closer inspection of the inferred pseudo-haplotypes associated with the swim component of the triathlon show that the *COL12A1* variant does not contribute to the interaction. Therefore, inferred pseudo-haplotypes were constructed from *COL3A1* rs1800255, *COL5A1* rs12722 and *COL6A1* rs35796750 and investigated for associations with the swim component. As expected, the G+C+T inferred pseudo-haplotype was significantly (p=0.038 and p=0.027) over-represented in both the Fast (19.2%, n=49) and Middle (17.8%, n=44) tertiles when compared to the Slow tertile (12.5%, n=29).

6.5.5 Multivariate Analysis for the Determination of Performance

A multivariate analysis was used to describe the relationship between participant characteristics and variant genotypes and the swim, bike and run components of the triathlon. Weight, and not BMI, was included in these analyses since height was not significantly correlated with overall race time. Therefore age, weight, race year and the independently associated variants were included in models to determine performance for the swim, bike and run components of the triathlon (Table 6.8). The variables race year and age predicted 37% of the variance ($p<0.001$, standard error of the estimate (SEE)=13.3) in the time to complete the swim component. Age, weight and the *COL6A1* rs35796750 genotype (TT vs. TC+CC) predicted 8% of the variance ($p<0.001$, SEE=39.7) in the bike component. For the run component, only weight predicted 15% of the variance ($p<0.001$, SEE=45.8).

Table 6.8. Multivariate analysis for time to complete the swim, bike and run components of the South African Ironman triathlon.

Multivariate analysis	***β***	***B***	**p value**
3.8km Swim time (min)			
Race Year	0.487	3.03	**<0.001**
Age (yrs)	0.276	0.57	**<0.001**
Weight (kg)	0.056	0.10	0.096
180km Bike time (min)			
Race Year	0.040	0.614	0.337
Age (yrs)	0.185	0.947	**<0.001**
Weight (kg)	0.160	0.708	**<0.001**
COL6A1 rs35796750 Genotype (TT vs TC+CC)	-0.093	-8.352	0.024
42.2km Run time (min)			
Race Year	0.045	0.83	0.259
Age (yrs)	0.067	0.42	0.095
Weight (kg)	0.376	2.01	**<0.001**
COL5A1 rs12722 Genotype (TT vs TC+CC)	0.054	5.58	0.171

For the 3.8km Swim time: R=0.611, adjusted R^2=0.370, SEE=13.3, p<0.0001. For the 180km Bike time: R=0.285, adjusted R^2=0.075, SEE=39.7, p<0.0001. For the 42.2km Run time: R=0.398, adjusted R^2=0.152, SEE=45.8, p<0.0001. p values in bold typeset are significant (p<0.05).

6.4 DISCUSSION

The first novel main finding of this study was the identification of a variant within the *COL6A1* gene as a performance marker for the bike component of the South African Ironman triathlon. The *COL6A1* rs35796750 C/T SNP was associated with performance during the 180km bike component of the 226km South African Ironman triathlon. In this study male triathletes with a TT genotype completed the bike of the triathlon significantly faster than those triathletes with a TC or CC genotype. In addition, a significant linear trend was found in the *COL6A1* rs35796750 TT genotype frequencies of the bike performance tertiles. Specifically the TT genotype was over-represented in the fastest bike performance tertile. Furthermore, the *COL6A1* rs35796750 genotype (TT vs TC+CC), age and weight contributed to 8% of the variance in the time to complete the bike of the South African Ironman triathlon. These results are in agreement with the hypothesis that the TT genotype would be beneficial in athletic performance, since the C allele is associated with negative outcomes in other multifactorial conditions. No significant independent associations were identified between this variant and time to complete the 3.8 km swim or 42.2 km run components of the events.

Interestingly, although not significant, there was a trend for triathletes with the *COL6A1* rs35796750 TT genotype to train for fewer hours when compared to those with TC or CC genotypes. Despite this, it may be concluded that this variant did not have a training effect since the TT genotype was associated with increased cycling performance. Further research is however required to test this hypothesis.

The second novel main finding of this study is that inferred pseudo-haplotypes constructed from the *COL3A1* rs1800255 (G/A), *COL6A1* rs35796750 (C/T) and *COL12A1* rs970547 (A/G) variants interact to modulate endurance running performance in four South African Ironman triathlons. The G+C+A inferred pseudo-haplotype, constructed from *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547, was over-represented in the Fast tertile when compared to the Slow tertile for the running component of the South African Ironman triathlon. Furthermore, since the *COL5A1* rs12722 variant was previously associated with endurance running [151], additional gene-gene interactions between *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 and endurance performance were investigated. Again, participants with the major G+T+C+A pseudo-haplotype were significantly over-represented in the Fast and Middle tertiles when compared to those in the Slow tertile when the running component of the triathlon was analysed. These findings implicate the *COL3A1*, *COL6A1* and *COL12A1* genes as novel potential markers for endurance running performance. Additional studies should investigate these genes in true endurance running events, such as marathons, in order to confirm the findings of this study. Furthermore, since no single variant independent associations were identified for *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 and endurance running performance, these findings highlight the importance of gene-gene interactions when investigating multi-genic complex phenotypes such as endurance performance.

Interestingly, gene-gene interactions were identified in this study, between *COL3A1* rs1800255, *COL5A1* rs12722 and *COL6A1* rs35796750, for time to complete the

swim component of the triathlon. A significant association was also identified between *COL3A1* rs1800255 and swimming training duration only. Specifically, participants with a GA genotype self-reported significantly more hours per week of swimming training, in the 15 weeks prior to the event, than their GG and AA counterparts. Mechanical loading is known to alter collagen gene expression however [92;93], since swimming is a non-load-bearing activity the reasons for these associations remain unknown and further research is needed to explain a possible mechanism.

Athletic performance is a multifactorial and polygenic phenotype [26;198]. Although this will be discussed in more detail in the final concluding chapter of this thesis (Chapter 10), it is important to note that a number of previous studies have associated genetic variants with athletic performance, specifically in the South African Ironman triathlon [43;151;172]. It must be stated that numerous extrinsic factors contribute to athletic performance as well [118]. It is therefore important that there were no differences in total distance (km/wk) or duration (hrs/wk) trained, or differences in distance (km/wk) or duration (hrs/wk) trained for each component of the triathlon, between the *COL6A1* rs35796750 or *COL12A1* rs970547 genotype groups among the participants with complete self-reported training history data. A significant association was identified between *COL3A1* rs1800255 and swimming training duration only. Specifically, participants with a GA genotype self-reported significantly more hours per week of swimming training, in the 15 weeks prior to the event, than their GG and AA counterparts. Mechanical loading is known to alter collagen gene expression however [92;93], since swimming is a non-load-bearing activity the reasons for this association remain unknown. No addition associations

were identified between *COL3A1* rs1800255 and self-reported training duration or training distance.

The lack of self-reported training history for the participants recruited during the 2000 and 2001 South African Ironman triathlon events is a limitation to this study. Training data for the 2000 and 2001 events would have allowed for the addition of self-reported training history to the multivariate analysis for times to complete each component of the triathlon. The inclusion of training history as well as other intrinsic and extrinsic factors may help explain more of the variance in each component than that explained by the current models. A second limitation to this study was that more South African born participants participated in the 2006 and 2007 events when compared to the 2000 and 2001 events. Therefore, although there may be effects due to population stratification, all participants were of self-reported Caucasian descent and there were no differences in weight, height or BMI between the four events. Furthermore, there were no genotype effects on country of birth. Finally, this study only investigated male triathletes. Gender is a known factor in the aetiology of athletic performance [26], and therefore additional studies are required to investigate these gene variants in female cohorts to determine if the associations identified in this study may be gender-specific, as shown when some of these variants were investigated in other exercise-associated phenotypes [137;153;154] (Chapter 3).

In conclusion this study identified, for the first time, an association between endurance cycling performance in the South African Ironman triathlon and a variant within the *COL6A1* gene. Specifically the *COL6A1* rs35796750 TT genotype is

associated with increased performance during the bike of the South African Ironman triathlon. In addition, a second novel finding of this study is that the *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 variants interact to modulate endurance running performance in four South African Ironman triathlons. Furthermore, these variants also interact with *COL5A1* rs12722 to modulate endurance running performance. These findings implicate collagen gene variants as novel potential markers of endurance running and cycling performance. Finally, a summary of the results of this chapter can be found in table 6.8.

Table 6.8. A continuing summary of the results of the chapters in this thesis. The results of chapter 6, and all preceding chapters, are shown.

Investigated Phenotype	Associations	Investigated Gene Variants			
		COL3A1 **rs1800255 (G/A)**	***COL5A1*** **rs12722 (C/T)**	***COL6A1*** **rs35796750 (C/T)**	***COL12A1*** **rs970547 (A/G)**
ACL	**Genotype**	not associated	CC ♀ prev pub[1]	not associated	AA ♀ prev pub[2]
	Haplotype	nd	T ♀[3]	nd	A ♀[3]
Achilles tendinopathy	**Genotype**	not associated prev pub[4]	CC prev pub[5]	not associated	not associated prev pub[6]
	Haplotype	A	C	C	nd
ROM	**Genotype**	not associated	CC prev pub[7]	not associated	not associated
	Haplotype	not associated	C	T	not associated
Endurance Swim Performance[9]	**Genotype**	not associated	not associated	not associated	not associated
	Haplotype	G	C	T	not associated
Endurance Cycle Performance[9]	**Genotype**	not associated	not associated	TT	not associated
	Haplotype	not associated	not associated	not associated	not associated
Endurance Running Performance[9]	**Genotype**	not associated	TT prev pub[8]	not associated	not associated
	Haplotype	G	T	C	A
EAMC		Chapter 7			
Rugby Union		Chapter 8			

Green shading indicates that the allele or genotype is associated with a reduced risk of injury, except for ROM and endurance performance where it represents an increased ROM or endurance performance. Red shading indicates that the allele or genotype is associated with an increased risk of injury. Grey shading indicates no independent association. Prev pub, previously published. ♀, associations identified only in females. ACL, anterior cruciate ligament; ROM, range of motion; EAMC, exercise-associated muscle cramps; nd, not determined.

[1] independent association previously reported by Posthumus et al. [153].
[2] independent association previously reported by Posthumus et al. [154], but not independently associated in this thesis.
[3] T+A pseudo-haplotype also associated with increased risk of female ACL ruptures when a Polish cohort was analysed [137].
[4] no independent association previously reported by Saunders [170]
[5] independent association previously reported by Mokone et al. [131] and repeated by September et al. [179].
[6] no independent association previously reported by September et al. [181]
[7] independent association previously reported by Brown et al. [28;29].
[8] independent association previously reported by Posthumus et al. [151] and repeated by Abrahams et al. [1].
[9] only male participants were included in this study.

CHAPTER 7

GENETIC RISK FACTORS FOR EXERCISE ASSOCIATED MUSCLE CRAMPS (EAMC)

The data presented in this chapter has been published in a condensed form in the peer-reviewed article: O'Connell K, Posthumus M, Schwellnus MP, Collins M. Collagen genes and exercise-associated muscle cramping. *Clinical Journal of Sport Medicine*. 2013;23(1):64-9.

7.1 INTRODUCTION

Exercise Associated Muscle Cramping (EAMC) is defined as “painful, spasmodic and involuntary contraction of skeletal muscle that occurs during or immediately after exercise” [176]. It is a common medical condition among participants of endurance events such as ultra-marathons and triathlons. EAMC has a reported lifetime prevalence of 26%, 30-50% and 67% in physical education students [134], marathon runners [125] and triathletes [88], respectively. Despite its high prevalence the aetiology of EAMC remains unknown [174]. A number of hypotheses to explain the primary factors contributing to EAMC have however been proposed. These include the electrolyte depletion, dehydration and the more recent altered neuromuscular control hypotheses [174].

As mentioned in chapter 2 section 2.5.4, the altered neuromuscular control hypothesis implicates the following primary factors in the development of EAMC i) an increased exercise intensity or duration [174;175;183], ii) the development of muscle

fatigue [174], iii) contraction of the muscle in a shortened position [174], and iv) possible tissue damage [174]. Pre-race serum creatine kinase activity showed a tendency to be higher in participants who developed EAMC during the 2009 56 km Two Oceans ultra-marathon when compared to 29 non-crampers [175]. Furthermore, cross-sectional studies have reported an association between a positive family history of EAMC and increased risk of EAMC [125;183]. A survey study showed an association between a family history of EAMC and a past history of EAMC in 1300 marathon runners [125]. This association was later confirmed in a case-control study of 433 Ironman triathletes where the frequency of a positive family history of EAMC was significantly more common in triathletes with a past history of EAMC when compared to those with no history of EAMC [183]. However, in two subsequent prospective cohort studies, a positive family history of EAMC was not associated with the risk of developing EAMC during an event in Ironman triathletes [178] and distance runners [175].

As established in chapter 2 and shown in previous experimental chapters within this thesis there is a genetic predisposition, specifically with collagen genes, to musculoskeletal soft tissue injuries (Chapters 3 and 4) [131;153;179] and athletic endurance performance (Chapter 6) [1;28;138;151]. Therefore, since muscle damage, increased exercise intensity and duration, and a familial predisposition may play a role in the aetiology of EAMC, it may be proposed that variants within collagen genes that code for components of the musculoskeletal system increase susceptibility to EAMC.

Mutations within the *COL3A1*, *COL5A1* and *COL12A1* genes are known to cause Ehlers-Danlos Syndrome (EDS) [33;217;218]. Although not a common symptom, muscle cramps have been reported as a clinical feature of EDS [168]. Bethlem myopathy [17;109;140] and Ullrich congenital muscular dystrophy [17;109;140] are caused by mutations within the *COL6A1* gene, highlighting its importance in muscle function. Since mutations in the *COL5A1*, *COL3A1*, *COL6A1* and *COL12A1* genes are known to result in connective tissue pathologies, it may be proposed that variants within these genes may modulate the risk of developing EAMC.

The aim of this study was therefore to determine if an association exists between past history of EAMC and *COL3A1* rs1800255 (G/A), *COL5A1* rs12722 (C/T), *COL6A1* rs35796750 (C/T) and *COL12A1* rs970547 (A/G) in participants of the South African Ironman triathlon and the Two Oceans ultra-marathon. Specifically, it was hypothesised that the *COL5A1* rs12722 TT, *COL6A1* rs35796750 TT and *COL12A1* rs970547 AA genotypes, and either *COL3A1* rs1800255 genotype, would be associated with increased risk of EAMC. Since the collagen genes might interact with one another to modulate the risk of musculoskeletal soft tissue injuries (Chapters 3 and 4) [71], the secondary aim was to investigate gene-gene interactions between the *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and/or *COL12A1* rs970547 variants and risk of EAMC. It was also hypothesised that the *COL5A1* rs12722 T, *COL6A1* rs35796750 T, *COL12A1* rs970547 A and either of the *COL3A1* rs1800255 alleles may be implicated in gene-gene interactions associated with increased risk of EAMC.

7.2 MATERIALS AND METHODS

7.2.1 Participants

Four hundred and forty-one male Caucasian participants were recruited at the registration of either the 2006 and 2007 226 km South African Ironman triathlon (n=271) or the 2009 and 2011 56 km Two Oceans ultra-marathon (n=170) for this genetic case-control association study (Figure 7.1). These events were chosen due to the high reported lifetime prevalence of EAMC in marathon runners and triathletes as described above. In addition only males were included in this study because previous studies have reported that males are more likely to experience EAMC [174;177]. Furthermore, as stated previously, athletic performance is associated with EAMC [174;177] and is also gender specific [26;198]. One hundred and eighteen participants reported a history of EAMC within the last 12 months prior to the event (EAMC group), while 150 reported no history of a previous (lifelong) EAMC (NON group). The participants included in this study therefore are not representative of all ultra-marathon runners or triathletes, but rather a selectively defined subset from these sports. Training data were only obtained during the Ironman events. Participants included in this study did not participate in both the Ironman and ultra-marathon events. Only the 2006 or 2009 data was analysed for those participants that were recruited at both Ironman or Two-Oceans events, respectively. This preference was due to the completeness of the respective data sets, as well as the specific data collected at each event.

All participants completed written informed consent forms (Appendix A.3) and a physical activity, medical history and cramping history questionnaire (Appendix A.4). Specifically, EAMC was defined as “painful, spontaneous, sustained spasm of a muscle during or immediately (within 6 hours) after exercise (in training or competition).” Approval for the study was obtained from the Human Research Ethics Committee of the University of Cape Town (Appendix A.1) and the organisers of the Ironman triathlon events.

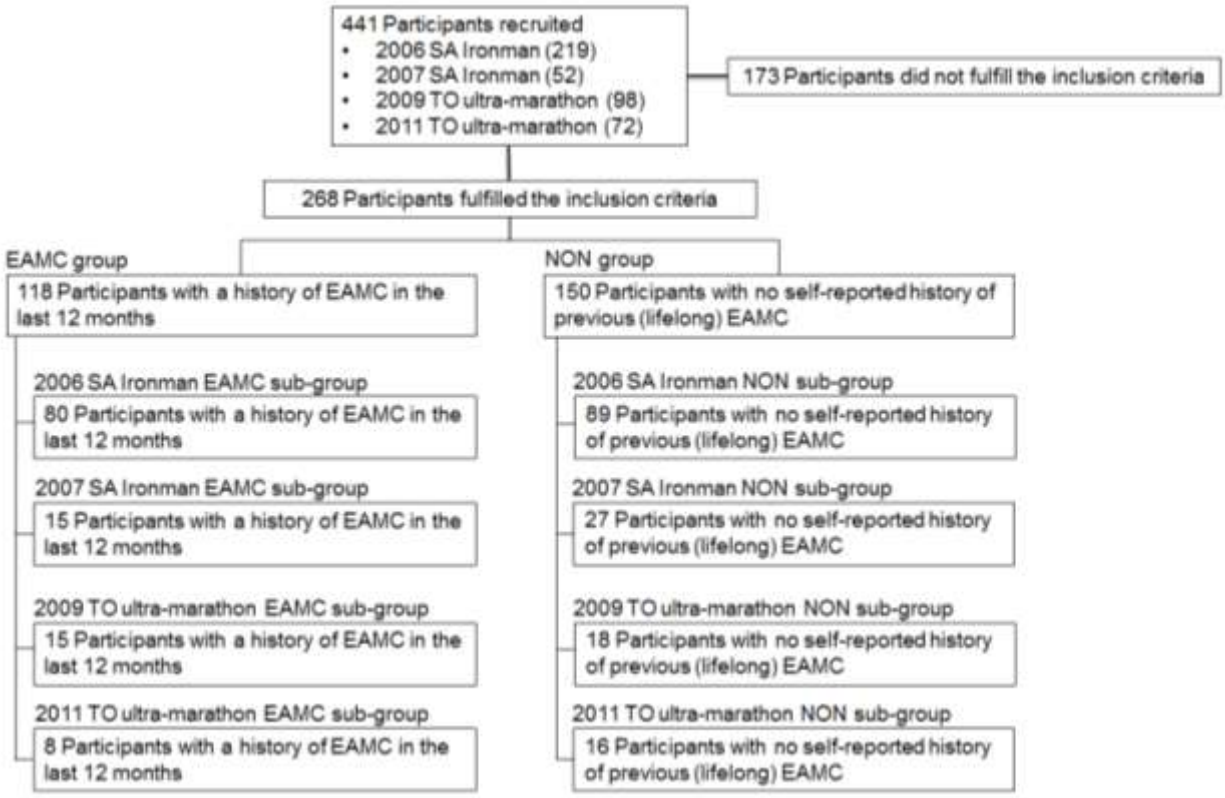


Figure 7.1. Flow diagram showing the total number of participants recruited at the South African Ironman triathlon and Two Oceans ultra-marathon events. In addition, the breakdown of included participants into those that reported a history of EAMC within 12 months prior to the event (EAMC group) and those with no self-reported history of previous (lifelong) EAMC (NON group), as well as the number of participants in each sub-group, are shown. EAMC, Exercise Associated Muscle Cramps. SA, South African. TO, Two Oceans.

7.2.2 Variant Selection

Four independent sequence variants were investigated in this study. Namely *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547. The reasons for their selection have been stated in chapter 3, section 3.2.2. Furthermore, these variants have been independently associated with, or interact to modulate, the risk of musculoskeletal soft tissue injuries (Chapter 3 and 4) and/or endurance performance (Chapter 6).

7.2.3 DNA Extraction and Genotyping Methods

Approximately 4.5ml of venous blood was collected from all participants by venipuncture of a forearm vein into an EDTA vacutainer tube. These samples were stored at 4˚C until DNA extraction, as previously described [108], with minor modifications [131]. An outline of this methodology was described in chapter 3, section 3.2.3.

All participants were genotyped for *COL5A1* rs12722, *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547. Genotyping of *COL3A1* rs1800255 and *COL6A1* rs35796750 was performed using custom designed Fluorescence-based Taqman® polymerase chain reaction (PCR) assays (Applied Biosystems, Foster City, CA, USA) as described in chapter 3. Genotyping of *COL5A1* rs12722 and *COL12A1* rs970547 was performed using PCR and restriction fragment length polymorphism analysis as previously described (Appendix B) [29;131;153;154] (Chapter 5). All genotyping was done at the UCT/MRC Research Unit for Exercise

Science and Sports Medicine, University of Cape Town, Cape Town, South Africa. All genotyping was done at the UCT/MRC Research Unit for Exercise Science and Sports Medicine, University of Cape Town, Cape Town, South Africa.

Investigators were blinded to the phenotypes of the participant samples while genotyping. Furthermore, a number of positive and negative controls were used to ensure genotyping accuracy. No discrepancies were observed. Of the 268 included participants, genotypes for *COL5A1* rs12722, *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 were obtained for 264 (98.5%), 228 (85.1%), 209 (78.0%) and 225 (84.0%) participants respectively. A number of positive and negative controls were used to ensure quality and accuracy of genotyping. No discrepancies were observed.

7.2.4 Statistical Analysis

The same statistical tests and haplotype analysis methods were performed as described in chapter 3 section 3.2.4. Age, weight, height, overall performance (where appropriate) and participant genotype data were used in a multivariate logistic regression analysis to determine their contribution to risk of past history of EAMC. Significance was accepted when $p<0.05$. No adjustments were made for multiple testing in this study for the same reasons as outlined in chapter 3 section 3.2.4.

7.3 RESULTS

7.3.1 Participant characteristics

Descriptive participant characteristics of the EAMC and NON groups are presented in Table 7.1. Age, height, weight, body mass index (BMI) and country of birth (% South African born) were similar between the EAMC and NON groups. Overall participant performance in the 2006 and 2007 Ironman triathlon and 2009 and 2011 Two Oceans ultra-marathon sub-groups are also presented in Table 7.1. The EAMC group (766.8 ± 94.0, n=95) was significantly faster (min) overall than the NON group (799.1 ± 95.8, n=116) in the 2006 and 2007 Ironman triathlon sub-group (p=0.015). There was no significant difference between the EAMC (295.0 ± 102.9, n=46) and NON (299.2 ± 106.8, n=25) groups for overall finishing time (min) in the 2009 and 2011 Two Oceans ultra-marathon sub-group (p=0.894). The distance (km/week) and duration (hrs/week) of training 15 weeks prior to the triathlon were not significantly different between the EAMC and NON groups (Table 7.2). Training data was not obtained from participants recruited at the 2009 and 2011 Two Oceans ultra-marathon events. There were no significant genotype effects on participant age, weight, height, BMI or country of birth (Table 7.3).

Table 7.1. General characteristics and finishing times of the participants that reported a history of EAMC within the last 12 months prior to the event (EAMC group) and the participants with a reported history of never suffering from EAMC (NON group).

	All (n=268)	**EAMC (n=118)**	**NON (n=150)**	**p-value**
Age (y)	38.4 ± 8.9 (266)	38.4 ± 8.9 (116)	38.4 ± 8.9 (150)	0.976
Height (cm)	179.4 ± 7.5 (245)	180.2 ± 7.3 (110)	178.7 ± 7.5 (135)	0.122
Weight (kg)	77.2 ± 10.7 (260)	77.9 ± 10.3 (115)	76.7 ± 11.0 (145)	0.360
BMI (kg/m^2)	23.9 ± 2.6 (244)	24.0 ± 2.4 (110)	23.9 ± 2.7 (134)	0.911
Country of Birth (% SA Born)	73.0 (143)	72.5 (66)	73.3 (77)	1.000
2006 and 2007 sub-groups Overall (min)	784.6 ± 96.1 (211)	766.8 ± 94.0 (95)	799.1 ± 95.8 (116)	0.015
2009 and 2011 sub-groups Overall (min)	297.3 ± 103.9 (46)	295.0 ± 102.9 (21)	299.2 ± 106.8 (25)	0.894

Values are expressed as means ± standard deviations, or percentages where appropriate. The number of participants is in parentheses. Numbers in bold typeset are significant ($p<0.05$) (EAMC vs. NON). BMI, Body Mass Index. SA, South Africa.

Table 7.2. Training history for the 2006 and 2007 Ironman triathlon sub-groups, during the 15 weeks prior to the triathlon, for the participants that reported a history of EAMC within the last 12 months prior to the event (EAMC group), and the participants with a reported history of never suffering from EAMC (NON group).

	All Participants [a] (n=211)	EAMC (n=95)	NON (n=116)	p value
Training duration (hrs/wk)				
3.8km Swim	3.1 ± 1.7 (203)	3.1 ± 1.9 (91)	3.1 ± 1.5 (112)	0.874
180km Bike	9.2 ± 12.8 (188)	10.3 ± 18.9 (84)	8.4 ± 2.8 (104)	0.328
42.2km Run	4.9 ± 3.1 (192)	5.3 ± 4.1 (86)	4.6 ± 1.8 (106)	0.122
226km Overall	15.9 ± 4.2 (177)	15.9 ± 4.0 (79)	15.8 ± 4.3 (98)	0.800
Training distance (km/wk)				
3.8km Swim	6.5 ± 3.0 (206)	6.8 ± 3.2 (93)	6.2 ± 2.8 (113)	0.164
180km Bike	226.3 ± 78.1 (188)	228.0 ± 74.9 (85)	224.9 ± 81.0 (103)	0.790
42.2km Run	47.6 ± 18.3 (200)	47.2 ± 14.6 (89)	48.0 ± 20.9 (111)	0.769
226km Overall	238.8 ± 77.9 (175)	243.8 ± 76.9 (81)	234.5 ± 78.9 (94)	0.430

Values are expressed as means ± standard deviations, or percentages where appropriate. The number of participants is in parentheses.

[a] Training history was not obtained from participants recruited at the 2009 and 2011 Two Oceans ultra-marathon events.

Table 7.3. Genotype effects of *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 on physiological characteristics of participants.

Polymorphism	p-values				
	Age	Height	Weight	BMI	Country of Birth
***COL3A1* rs1800255**	0.923	0.430	0.643	0.641	0.916
***COL5A1* rs12722**	0.285	0.495	0.743	0.827	0.471
***COL6A1* rs35796750**	0.433	0.695	0.447	0.458	0.398
***COL12A1* rs970547**	0.596	0.253	0.486	0.307	0.806

p-values in bold typeset indicate significant differences ($p<0.05$). BMI, body mass index.

7.3.2 Nature of the self-reported past history of EAMC

A trend was observed for the frequency of self-reported previous (life-long) tendon or ligament injury to be higher in participants with a past history of EAMC ($p=0.050$). Specifically, the frequency of self-reported injuries was lower in the NON group (33.8%) when compared to the EAMC group (46.4%).

7.3.3 Independent Collagen Gene Associations and History of Exercise Associated Muscle Cramping

COL5A1 rs12722 genotype frequency was significantly associated with past history of EAMC ($p=0.034$) (Figure 7.2B). Specifically the *COL5A1* rs12722 CC genotype was significantly under-represented ($p=0.031$, OR=2.2, 95% CI 1.1–4.5) in the EAMC group (CC genotype 11.1%) when compared to the NON group (CC genotype

21.8%). There were no significant differences in the *COL3A1* rs1800255 (Figure 7.2A; p=0.828), *COL6A1* rs35796750 (Figure 7.2C; p=0.300) or *COL12A1* rs970547 (Figure 7.2D; p=0.120) genotype frequencies between the EAMC and NON groups. The *COL12A1* rs970547 EAMC group was out of Hardy Weinberg equilibrium (p=0.011). The EAMC and NON groups for *COL3A1* rs1800255 (EAMC p=0.627, NON p=1.000), *COL5A1* rs12722 (EAMC p=1.000, NON p=0.324) and *COL6A1* rs35796750 (EAMC p=1.000, NON p=0.723), as well as, the *COL12A1* rs970547 NON group (p=0.309) were in Hardy Weinberg equilibrium.

Participants in the EAMC group were divided in three sub-groups based on their number of self-reported cramps suffered in the last ten races. Participants reported suffering from no cramps, 1 cramp or 2+ cramps in the last ten races. No genotype effect was determined for *COL5A1* rs12722 (p=0.480), *COL3A1* rs1800255 (p=0.112), *COL6A1* rs35796750 (p=0.181) or *COL12A1* rs970547 (p=0.759) and number of self-reported cramps in the last ten races.

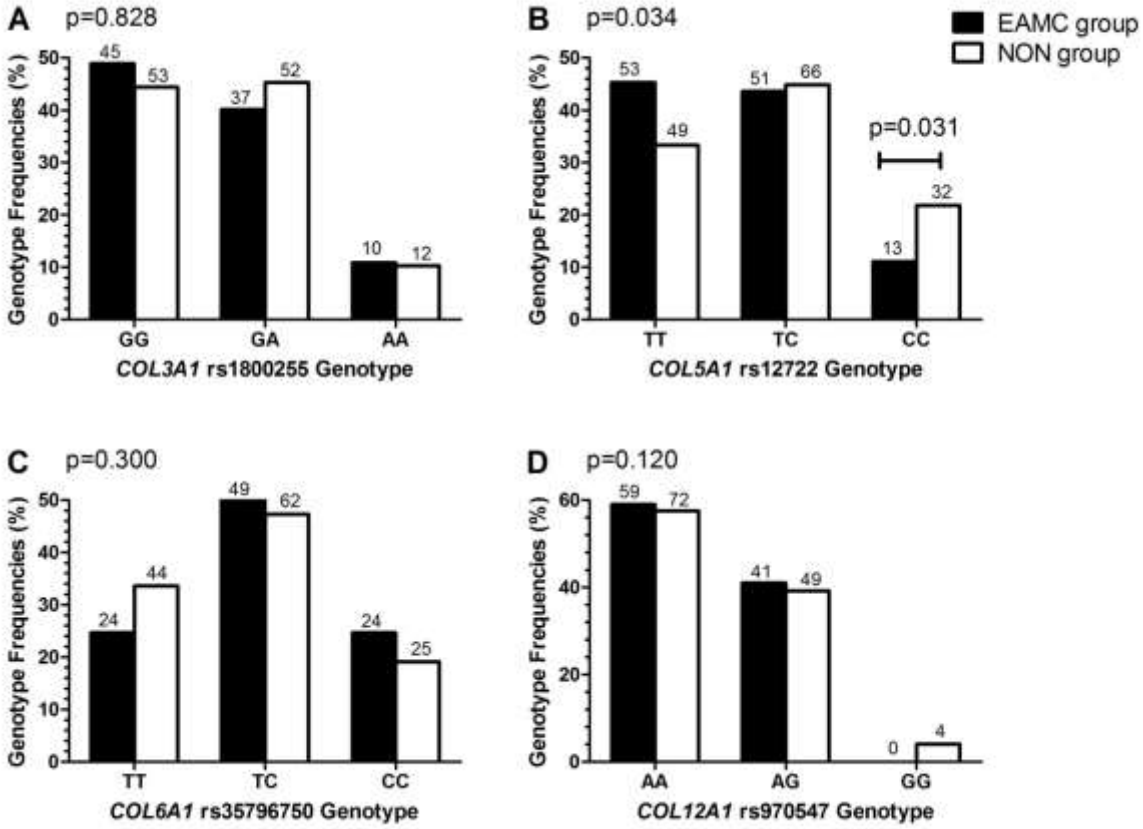


Figure 7.2. Genotype frequencies for the participants that reported a history of EAMC within 12 months prior to the event (EAMC group) and those with no self-reported history of previous (lifelong) EAMC (NON group) for the (A) *COL3A1* rs1800255, (B) *COL5A1* rs12722, (C) *COL6A1* rs35796750 and (D) *COL12A1* rs970547 polymorphisms. Numbers of participants (n) are indicated above each specific column. Values above the figures refer to the overall p-value. p-values above a genotype group refer to the pairwise post-hoc analysis.

7.3.4 Gene-Gene Interactions and History of Exercise Associated Muscle Cramping

Gene-gene interactions between *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 were investigated (Figure 7.3). The major G+T+C+A inferred pseudo-haplotype was significantly (p=0.028) over-represented in the EAMC group (22.7%, n=27) when compared to the NON group (14.2, n=21).

Furthermore, the G+C+T+A inferred pseudo-haplotype was significantly (p=0.021) under-represented in the EAMC group (11.1%, n=13) when compared to the NON group (16.1%, n=24) (Figure 7.3).

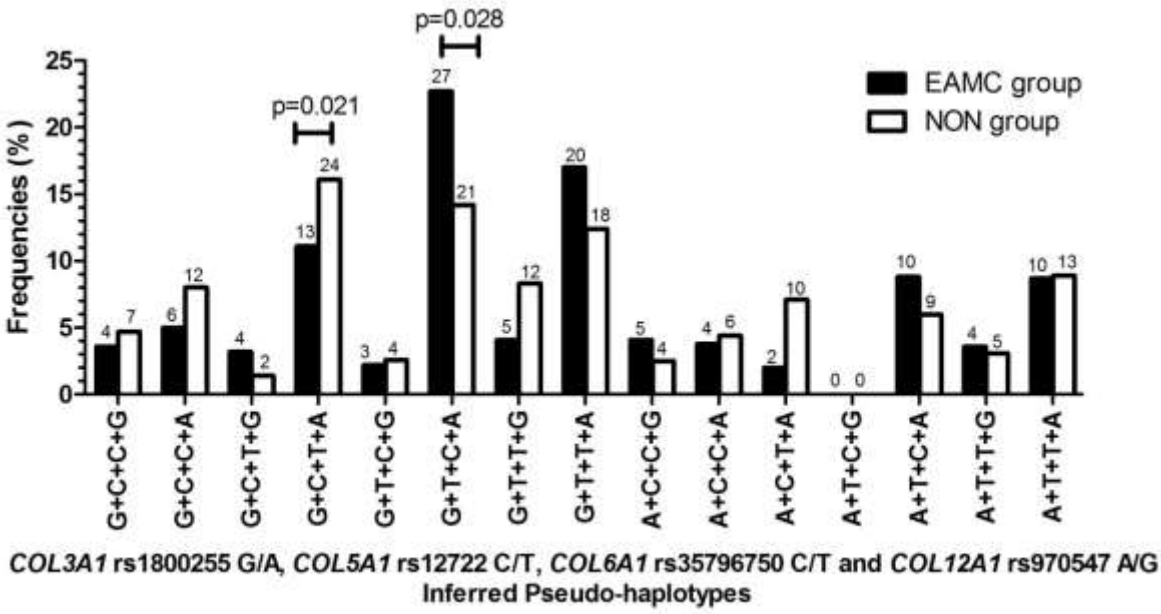


Figure 7.3. Frequency distributions of the inferred pseudo-haplotypes constructed from *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 between participants that reported a history of EAMC within 12 months prior to the event (EAMC group) and those with no self-reported history of previous (lifelong) EAMC (NON group).

Among the two significantly identified inferred pseudo-haplotypes the direction of association for the *COL5A1* rs12722 and *COL6A1* rs35796750 variants was consistent, while *COL3A1* rs1800255 and *COL12A1* rs970547 showed no evidence of contributing to the inferred pseudo-haplotypes (Figure 7.3). Therefore additional inferred pseudo-haplotypes were constructed from only *COL5A1* rs12722 T/C and

COL6A1 rs35796750 T/C. The T+C inferred pseudo-haplotype was significantly (p=0.011) over-represented in the EAMC group (33.1%, n=39) when compared to the NON group (23.9%, n=35) (Figure 7.4). Furthermore, the C+T pseudo-haplotype was significantly (p=0.007) under-represented in the EAMC group (16.0 %, n=19) when compared to the NON group (25.3%, n=37) (Figure 7.4).

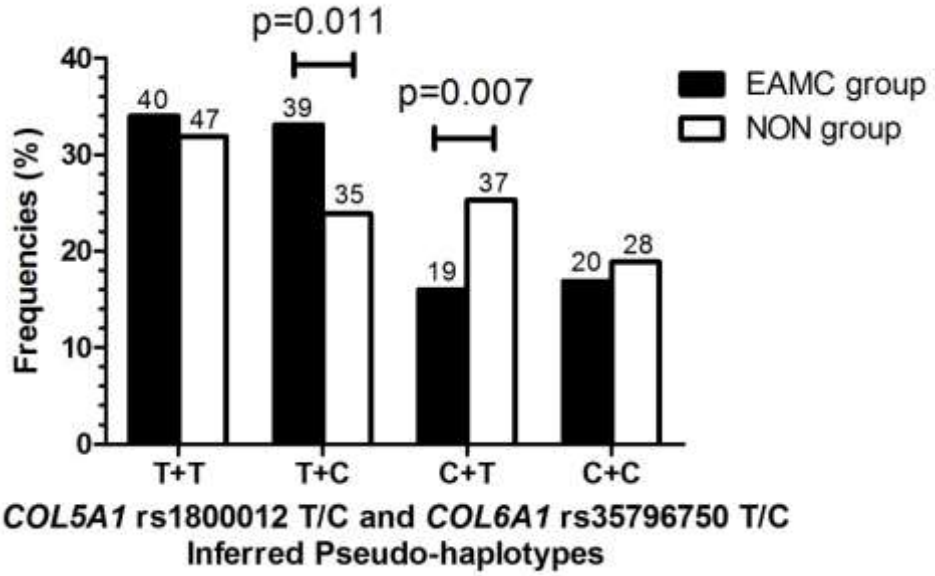


Figure 7.4. Frequency distributions of the inferred pseudo-haplotypes constructed from *COL5A1* rs12722 and *COL6A1* rs35796750 between participants that reported a history of EAMC within 12 months prior to the event (EAMC group) and those with no self-reported history of previous (lifelong) EAMC (NON group).

7.3.5 Logistic Regression Analysis for Risk of Past History of EAMC

Participant weight and height have previously been identified as risk factors for past history of EAMC [29]. Furthermore, the *COL5A1* rs12722 CC genotype is

independently associated with past history of EAMC (Figure 7.2A) and self-reported history of tendon or ligament injury showed a strong trend with risk of past history of EAMC. Therefore, participant weight (kg), height (cm), self-reported history of tendon or ligament injury (Yes vs. No) and *COL5A1* rs12722 genotype (CC vs. TC+TT) were incorporated into a logistic regression model to determine risk of past history of EAMC in all participants (Table 7.4). The *COL5A1* rs12722 genotype (CC vs. TC+TT) (OR=3.09, 95% CI = 2.64-3.54, p=0.013) and self-reported history of tendon or ligament injury (Yes vs. No) (OR=2.03, 95% CI = 1.72-2.34, p=0.024) significantly contributed to past history of EAMC.

Table 7.4. Regression analysis for the determination of risk of past history of EAMC.

	Level of effect	OR (95% CI)	Wald Statistic	p value
Weight (kg)	-	1.00 (0.96-1.04)	-0.036	0.998
Height (cm)	-	1.05 (0.99-1.10)	2.772	0.096
Tendon/Ligament Injury (Yes vs. No)	Yes	2.03 (1.72-2.34)	5.094	**0.024**
***COL5A1* rs12722 genotype (CC vs. TC+TT)**	TC+TT	3.09 (2.64-3.54)	6.113	**0.013**

Numbers in bold typeset are significant contributors to the model for past history of EAMC ($p<0.05$). OR, odds ratio. CI, confidence interval.

Furthermore, in order to determine the effect of overall performance on a past history of EAMC, a separate logistic regression was also determined analysing only the participants from the 2006 and 2007 Ironman triathlon sub-groups (Table 7.5). As before, participant weight (kg), height (cm), self-reported history of tendon or ligament injury (Yes vs. No) and *COL5A1* rs12722 genotype (CC vs. TC+TT) were

incorporated. Furthermore, overall finishing time for the Ironman triathlon was also included in the model since the EAMC group was shown to be significantly faster than the NON group (Table 7.1). Training history was not included in the model since no significant interactions were observed (Table 7.2). The *COL5A1* rs12722 genotype (CC vs. TC+TT) (OR=2.85, 95% CI = 2.40-3.30, p=0.022), self-reported history of tendon or ligament injury (Yes vs. No) (OR=1.98, 95% CI = 1.67-2.29, p=0.031) and overall finishing time for the Ironman triathlon (OR=1.00, 95% CI = 1.00-1.00, p=0.047) significantly contributed to past history of EAMC.

Table 7.5. Regression analysis for the determination of risk of past history of EAMC in participants of the 2006 and 2007 South African Ironman triathlons.

	Level of effect	OR (95% CI)	Wald Statistic	p value
Weight (kg)	-	1.02 (0.98-1.06)	0.859	0.354
Height (cm)	-	1.03 (0.97-1.09)	1.085	0.298
Overall performance (min)	-	1.00 (1.00-1.00)	3.955	**0.047**
Tendon/Ligament Injury (Yes vs. No)	Yes	1.98 (1.67-2.29)	4.651	**0.031**
***COL5A1* rs12722 genotype (CC vs. TC+TT)**	TC+TT	2.85 (2.40-3.30)	5.207	**0.022**

Numbers in bold typeset are significant contributors to the model for past history of EAMC ($p<0.05$). SE, standard error of the estimate.

7.4 DISCUSSION

This is the first study to identify a genetic variant as a potential marker for a past history of exercise-associated muscle cramps. The *COL5A1* 3'-UTR rs12722 (C/T) polymorphism was associated with self-reported EAMC during the 12 months prior to participating in either the 2006 and 2007 South African Ironman triathlons or the 2009 and 2011 Two Oceans ultra-marathons. Specifically, the CC genotype was significantly over-represented in participants with no self-reported history of previous (lifelong) EAMC when compared to participants that reported a past history of EAMC within 12 months prior to the event. In addition, a significant interaction between *COL5A1* rs12722 and *COL6A1* rs35796750 was identified. Specifically, the T+C pseudo-haplotype, constructed from *COL5A1* rs12722 T/C and *COL6A1* rs35796750 T/C, was associated with increased risk of a past history of EAMC, while the C+T pseudo-haplotype was significantly associated with reduced risk. Multivariate regression analysis of all participants revealed the *COL5A1* rs12722 genotype (TC+TT) and previous tendon/ligament injury as positive contributors to risk of past history of EAMC. Furthermore, when only participants of the 2006 and 2007 Ironman triathlons sub-group were analysed, the *COL5A1* rs12722 genotype (TC+TT), previous tendon/ligament injury and faster overall finishing time were all positive contributors to risk of past history of EAMC. No independent associations were identified between *COL3A1* rs1800255, *COL6A1* rs35796750 or *COL12A1* rs970547 and risk of past history of EAMC.

This novel study shows that structural components (types V and VI collagen) of the extracellular matrix of musculoskeletal soft tissue are associated with a past history

of EAMC. Prospective cohort studies are required to establish a cause-effect relationship however, the *COL5A1* genetic continuum discussed in chapter 2 section 2.2 suggests a possible molecular mechanism [107]. Specifically, the *COL5A1* 3'-UTR functional form containing the rs12722 T allele may contribute to increase mRNA stability [107]. These changes in *COL5A1* expression may alter the structural and mechanical properties of collagen fibrils [41]. It has been proposed that the increased type V collagen production associated with the TT genotype of rs12722 results in increased stiffness and/or creep inhibition, and reduced tensile strength [41]. These changes increase the risk of musculoskeletal soft tissue injuries. In contrast, the CC genotype is proposed to have reduced stiffness and greater tensile strength [41]. The changes in these mechanical properties may explain the reduced risk of EAMC within individuals with a CC genotype. Further work is required to test this hypothesis. The function of type VI collagen is still largely unknown [201], however type V and type VI collagen are known to interact [190]. Therefore, although a function for *COL6A1* rs35796750 has been proposed [191], additional studies are required to determine the functional effects of this variant to add to the understanding of how the *COL5A1* and *COL6A1* genes may interact to modulate the risk of EAMC.

In addition to the association identified with *COL5A1* rs12722, and the interaction with *COL6A1* rs35796750, a faster overall finishing time was also associated with a past history of EAMC in the 2006 and 2007 Ironman triathlon events. This result was expected since increased exercise performance has previously been associated with risk of EAMC in larger prospective cohorts from the 2006 and 2007 South African Ironman triathlons [183] and the 2009 Two Oceans ultra-marathon [175]. The

COL5A1 rs12722 TT genotype is also associated with increased endurance running performance [151]. Interestingly, multivariate logistic analysis revealed that both the *COL5A1* rs12722 T allele and a faster overall finishing time contribute to an increased risk of past history of EAMC. Future studies are required to further investigate this relationship between exercise performance, EAMC and *COL5A1* rs12722. In addition, the complex multifactorial aetiology of EAMC will be discussed in more detail in the final concluding chapter of this thesis (Chapter 10).

EAMC is a multifactorial phenotype [174] and therefore, although this will be discussed in more detail in the final concluding chapter of this thesis (Chapter 10), one of the main limitations of this study is that the regression analyses only identified overall finishing time, history of tendon or ligament injuries (Yes vs. No) and *COL5A1* rs12722 genotype (CC vs. TC+TT) as significant contributors to risk of past history of EAMC. Logistic estimates revealed the *COL5A1* rs12722 TT and TC genotypes, a history of tendon or ligament injuries and a faster overall finishing time as positive contributors to risk of past history of EAMC, however other previously described risk factors [183] such as weight and height an did not contribute to the model. This highlights the need for replication and validation of these associations in other larger and independent cohorts. In addition, these results demonstrate the importance of identifying new risk factors that may help to further explain the aetiology of EAMC. A second major limitation to this study is that muscle damage is only proposed to alter risk of EAMC [174] and that creatine kinase activity is weakly associated with risk of EAMC [175]. In order to add weight to the hypothesis that alterations to the collagen fibril mechanical properties as a result of *COL5A1* rs12722, or other collagen gene variants, modulate the risk of EAMC, further studies are required to investigate the

relationship between risk of EAMC and muscle damage. Additional limitations to this study include a small sample size and the use of self-reported data, such as past history of EAMC. In order to strengthen these findings additional studies are required in larger independent cohorts containing individuals diagnosed with EAMC. Finally, only males were included in this study because previous studies have reported that males are more likely to experience EAMC [174;177]. Furthermore, as stated previously, athletic performance is associated with EAMC [174;177] and is also gender specific [26;198]. Additional studies are therefore required to investigate these gene variants in female cohorts to determine if the associations identified in this study may be gender-specific, as shown when some of these variants were investigated in other exercise-associated phenotypes [137;153;154] (Chapter 3).

In conclusion, the main finding of this study was the novel identification of *COL5A1* rs12722 as a potential marker for a past history of EAMC. Specifically the *COL5A1* rs12722 CC genotype was significantly over-represented in participants that reported never suffering from EAMC, when compared to participants that reported suffering from EAMC within 12 months prior to the events. A significant interaction was also identified between the *COL5A1* and *COL6A1* genes and risk of past history of EAMC. These results suggest that changes to collagen containing connective tissue may directly and/or indirectly modulate the risk of a past history of EAMC, however further investigation is required. A summary of the results of this chapter are presented in table 7.6.

Table 7.6. A continuing summary of the results of the chapters in this thesis. The results of chapter 7, and all preceding chapters, are shown.

Investigated Phenotype	Associations	Investigated Gene Variants			
		COL3A1 **rs1800255 (G/A)**	***COL5A1*** **rs12722 (C/T)**	***COL6A1*** **rs35796750 (C/T)**	***COL12A1*** **rs970547 (A/G)**
ACL	**Genotype**	not associated	CC ♀ prev pub[1]	not associated	AA ♀ prev pub[2]
	Haplotype	nd	T ♀[3]	nd	A ♀[3]
Achilles tendinopathy	**Genotype**	not associated prev pub[4]	CC prev pub[5]	not associated	not associated prev pub[6]
	Haplotype	A	C	C	nd
ROM	**Genotype**	not associated	CC prev pub[7]	not associated	not associated
	Haplotype	not associated	C	T	not associated
Endurance Swim Performance[9]	**Genotype**	not associated	not associated	not associated	not associated
	Haplotype	G	C	T	not associated
Endurance Cycle Performance[9]	**Genotype**	not associated	not associated	TT	not associated
	Haplotype	not associated	not associated	not associated	not associated
Endurance Running Performance[9]	**Genotype**	not associated	TT prev pub[8]	not associated	not associated
	Haplotype	G	T	C	A
EAMC[9]	**Genotype**	not associated	CC	not associated	not associated
	Haplotype	not associated	C	T	not associated
Rugby Union		Chapter 8			

Green shading indicates that the allele or genotype is associated with a reduced risk of injury, except for ROM and endurance performance where it represents an increased ROM or endurance performance. Red shading indicates that the allele or genotype is associated with an increased risk of injury. Grey shading indicates no independent association. Prev pub, previously published. ♀, associations identified only in females. ACL, anterior cruciate ligament; ROM, range of motion; EAMC, exercise-associated muscle cramps; nd, not determined.

[1] independent association previously reported by Posthumus et al. [153].
[2] independent association previously reported by Posthumus et al. [154], but not independently associated in this thesis.
[3] T+A pseudo-haplotype also associated with increased risk of female ACL ruptures when a Polish cohort was analysed [137].
[4] no independent association previously reported by Saunders [170]
[5] independent association previously reported by Mokone et al. [131] and repeated by September et al. [179].
[6] no independent association previously reported by September et al. [181]
[7] independent association previously reported by Brown et al. [28;29].
[8] independent association previously reported by Posthumus et al. [151] and repeated by Abrahams et al. [1].
[9] only male participants were included in this study.

CHAPTER 8

GENETIC MARKERS FOR RUGBY UNION PLAYING LEVEL AND POSITION

8.1 INTRODUCTION

Rugby union football, hereafter referred to as rugby, is a global recreational and professional team sport which involves both high (examples: tackling, scrummaging and sprinting) and low (examples: standing, walking and jogging) intensity activities [167]. Teams are comprised of 15 players and can be divided into two discrete sub-groups; forwards (playing positions 1-8) and backs (playing positions 9-15) [167]. Although it is expected that all 15 players, irrespective of their position or sub-group, are competent at all aspects of the game, a high level of specialisation still remains within and between these sub-groups [167]. Within a single game of rugby, backs have been shown to cover a greater total distance than forwards [167]. Furthermore, within the backs sub-group, outside backs (playing positions 11, 14 and 15) sprinted greater distances and covered less distance at medium-intensity running or while jogging when compared to inside backs (playing positions 9, 10, 12 and 13) [167].

A number of studies have reported differences in player characteristics between playing positions and playing level [157;158;185]. However, these descriptive studies do not help to determine the underlying factors which may be over-represented in

individuals within particular playing positions or playing levels. Only two studies have investigated genetic markers for determining individual playing position [14;15] and playing level [14]. The first of these studies showed that there were differences in measures of leg power between forwards and backs when the players were grouped according to their *ACE* rs5186 (I/D) genotypes [15]. A second study by the same group investigated the functional *ACTN3* rs1815739 (R/X) gene variant as a marker of playing position and level [14]. No associations were identified between *ACTN3* rs1815739 and playing position or playing level [14].

To date, a number of collagen gene variants have been implicated in the aetiology of exercise-related phenotypes [28;29;40;71;131;138;139;151;153-155;171;179]. As previously discussed *COL5A1* rs12722 has been associated with range of motion [28;29;40] and endurance running performance [28;151]. Furthermore the functional *COL5A1* 3'-UTR region [107] , containing rs12722, may affect levels of the type V collagen α1 chain synthesis and thereby modulate fibrillogenesis and the mechanical properties of connective tissues [41]. Therefore, it was hypothesised that the *COL5A1* gene may be a good candidate for associations with other endurance or power based sports in which musculoskeletal soft tissue mechanical properties may be regarded as an important feature, such as rugby union [41]. Fibrillogenesis is also regulated by, amongst other proteins, the three other collagens, namely types III, VI and XII, which are the focus of this thesis [56;78;190;213]. Interestingly, types VI and XII collagen are also known to interact with type V collagen within the fibril [78;190;213]. The *COL6A1* rs35796750 variant is independently implicated in endurance cycling performance (Chapter 6) [138]. In addition, inferred pseudo-haplotypes constructed from *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1*

rs35796750 and *COL12A1* rs970547 were implicated in endurance running performance (Chapter 6).

Therefore the primary aim of this study was to determine if *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 are associated with playing level in rugby union. The secondary aim was to investigate these variants for associations within and between the forwards and backs playing position sub-groups. The third aim was to investigate gene-gene interactions between *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and/or *COL12A1* rs970547 for associations with playing level in rugby union and within and between the forwards and backs playing positions sub-groups.

8.2 MATERIALS AND METHODS

8.2.1 Participants

Three hundred and fifty-three self-reportedly Caucasian unrelated participants were recruited for this case-control genetic association study. One hundred and seventy professional rugby union players were recruited from South African teams competing the in the Super 15 rugby tournament (rugby group). The rugby group was further sub-divided into two sub-groups; forwards (playing positions 1-8, n=103) and backs (playing positions 9-15, n=67). An additional 183 participants, made up of 85 participants that had never played rugby and 98 recreational rugby players, were recruited as controls (control group). Rugby union is played predominantly by males and so this study included only male participants.

Prior to participation in this study, all the participants gave informed written consent (Appendix A.3). In addition, each participant completed a sports participation questionnaire (Appendix A.4). This study was approved by the Human Research Ethics Committee of the Faculty of Health Sciences within the University of Cape Town, South Africa (Appendix A.1).

8.2.2 Variant Selection

Four independent sequence variants were investigated in this study. Namely *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547. The reasons for their selection have been stated in chapter 3, section 3.2.2. Furthermore, these variants have been independently associated with, or interact to modulate, the risk of musculoskeletal soft tissue injuries (Chapter 3, 4 and 7) and/or endurance performance (Chapter 6).

8.2.3 DNA Extraction and Genotyping Methods

Approximately 4.5ml of venous blood was collected from all participants by venipuncture of a forearm vein into an EDTA vacutainer tube. These samples were stored at 4˚C until DNA extraction, as previously described [108], with minor modifications [131]. An outline of this methodology was described in chapter 3, section 3.2.3.

All participants were genotyped for *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547. Genotyping of *COL3A1* rs1800255 and

COL6A1 rs35796750 was performed using custom designed Fluorescence-based Taqman® polymerase chain reaction (PCR) assays (Applied Biosystems, Foster City, CA, USA) as described in chapter 3. Genotyping of *COL5A1* rs12722 and *COL12A1* rs970547 was performed using PCR and restriction fragment length polymorphism analysis as previously described (Appendix B) [29;131;153;154] (Chapter 5). All genotyping was done at the UCT/MRC Research Unit for Exercise Science and Sports Medicine, University of Cape Town, Cape Town, South Africa. All genotyping was done at the UCT/MRC Research Unit for Exercise Science and Sports Medicine, University of Cape Town, Cape Town, South Africa.

Investigators were blinded to the phenotypes of the participant samples while genotyping. Furthermore, a number of positive and negative controls were used to ensure genotyping accuracy. No discrepancies were observed. Of the 353 included participants, a total of 301 (85.3%), 312 (88.4%), 310 (87.8%) and 317 (89.8%) participants were genotyped for *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 respectively. All laboratory work was conducted at the MRC/UCT Research Unit for Exercise Science & Sports Medicine, University of Cape Town, Cape Town, South Africa.

8.2.4 Statistical Analysis

The same statistical tests and haplotype analysis methods were performed as described in chapter 3 section 3.2.4. Significance was accepted when $p<0.05$. No adjustments were made for multiple testing in this study for the same reasons as outlined in chapter 3 section 3.2.4.

8.3 RESULTS

8.3.1 Participant Characteristics

The genotype frequency distributions between the participants that had never played rugby and the recreational rugby players were similar for *COL3A1* rs1800255 (p=0.988), *COL5A1* rs12722 (p=0.152), *COL6A1* rs35796750 (p=0.469) and *COL12A1* rs970547 (p=0.211), and were therefore these participants were combined for all further analysis as a single CON group (Appendix C.10). The general physical characteristics for all participants are presented in table 8.1. The control group was significantly older than the rugby group. The rugby group was significantly taller and when compared to the control group. Similar results were identified for the forwards and backs sub-groups, except that the backs and control group were matched for height (Table 8.1). The forwards sub-group was significantly taller ($p<0.001$), heavier ($p<0.001$), with a corresponding higher BMI ($p<0.001$) than the backs sub-group. Furthermore, approximately 28% (n=48) of the rugby group played Rugby Union at an international level.

Although probably not biologically relevant, participants with the *COL3A1* rs1800255 GA genotype (186.1 ± 7.9 cm, n=130) were significantly ($p<0.001$) taller than those with the GG (182.5 ± 6.6 cm, n=143) or AA (183.0 ± 8.1 cm, n=20) genotypes. There were no additional genotype effects on any of the participant characteristics for *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 (Table 8.2).

Table 8.1. Physical characteristics for the professional rugby union players (rugby group) and participants that had never played rugby or played recreational rugby (control group), as well as the forwards and backs sub-groups.

	Control Group (183)	Rugby Group (170)	Forwards Sub-group (103)	Backs Sub-group (67)	p value[a]	p value[b]	p value[c]
Age (yrs)	30.7 ± 9.5 (178)	28.1 ± 5.3 (166)	28.1 ± 5.5 (99)	28.0 ± 5.0 (67)	**0.002**	**0.016**	**<0.001**
Height (cm)	181.5 ± 6.7 (175)	186.6 ± 7.6 (168)	190.0 ± 6.7 (101)	181.4 ± 5.7 (67)	**<0.001**	**<0.001**	0.970
Weight (kg)	82.8 ± 11.2 (178)	102.5 ± 12.9 (170)	110.8 ± 8.6 (103)	89.8 ± 6.2 (67)	**<0.001[d]**	**<0.001[d]**	**<0.001[d]**
BMI (kg.m-2)	24.9 ± 4.0 (173)	29.4 ± 3.0 (168)	30.7 ± 3.0 (101)	27.3 ± 1.3 (67)	**<0.001[d]**	**<0.001[d]**	**<0.001[d]**

Values are expressed as means ± standard deviations. The number of participants is in parentheses. Numbers in bold typeset are significant ($p<0.05$). BMI, Body Mass Index.

[a] rugby group vs. control group.

[b] forwards sub-group vs. control group.

[c] backs sub-group vs. control group.

[d] Co-varied for age.

Table 8.2. Genotype effects of *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 on physiological characteristics of participants.

Polymorphism	**p-values**		
	Age	**Height**	**Weight**
***COL3A1* rs1800255**	0.189	**<0.001**	0.074
***COL5A1* rs12722**	0.582	0.945	0.935
***COL6A1* rs35796750**	0.932	0.808	0.735
***COL12A1* rs970547**	0.778	0.369	0.935

Numbers in bold typeset are significant ($p<0.05$).

8.3.2 Collagen Genes and Rugby Union Playing Level and Position

There were no differences in genotype frequency distributions, for *COL3A1* rs1800255 (p=0.275), *COL5A1* rs12722 (p=0.941), *COL6A1* rs35796750 (p=0.883) or *COL12A1* rs970547 (p=0.447), between the control and rugby groups (Figure 8.1). When the forwards and backs sub-groups were analysed the *COL3A1* rs1800255 variant was associated with rugby union playing position (Figure 8.1A). Specifically, the *COL3A1* rs1800255 GG genotype was significantly (p=0.047) under-represented in the forwards sub-group (GG genotype: 39.6%, n=38) when compared to the control group (GG genotype: 53.5%, n=76) (Figure 8.1A). Furthermore, the *COL3A1* rs1800255 GA genotype was significantly (p=0.046) over-represented in the forwards sub-group (GA genotype: 53.1%, n=51) when compared to the control group (GA genotype: 39.4%, n=56) (Figure 8.1A). No other significantly different genotype frequency distributions, for *COL3A1* rs1800255 (backs sub-group p=0.962), *COL5A1* rs12722 (forwards sub-group p=0.782, backs sub-group p=0.617), *COL6A1* rs35796750 (forwards sub-group p=0.974, backs sub-group p=0.790) or *COL12A1* rs970547 (forwards sub-group p=0.772, backs sub-group p=0.102), were identified between the forwards or backs sub-groups and the control group (Figure 8.1).

Interestingly, when the smaller outside backs sub-group (playing positions 11, 14 and 15, n=18) was analysed the *COL12A1* rs970547 GG genotype was significantly over-represented in the outside backs sub-group (GG genotype: 23.5%, n=4) when compared to the control group (GG genotype: 5.8%, n=9, p=0.028), forwards sub-

group (GG genotype: 8.6%, n=5, p=0.026) and all other positions (GG genotype: 6.8%, n=10, p=0.043), respectively.

The *COL3A1* rs1800255 (rugby group p=0.141, control group p=1.000), *COL5A1* rs12722 (rugby group p=0.744, control group p=1.000), *COL6A1* rs35796750 (rugby group p=1.000, control group p=0.742) and *COL12A1* rs970547 (rugby group p=1.000, control group p=0.834) variants were in Hardy-Weinberg equilibrium.

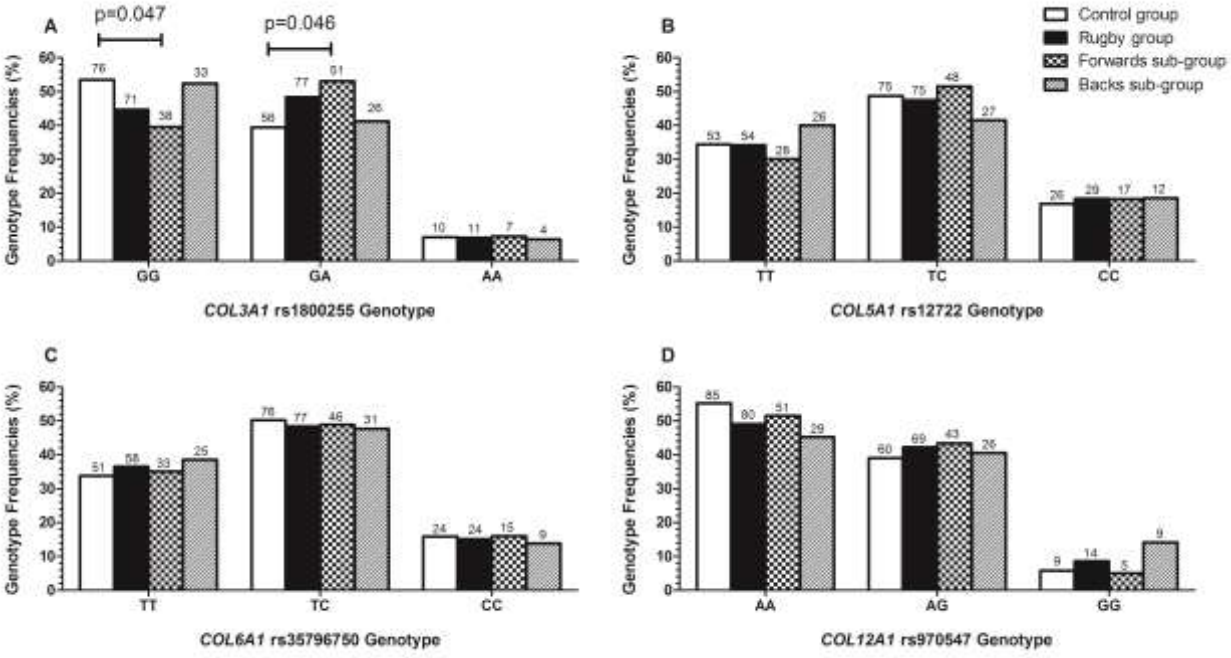


Figure 8.1. Genotype frequency distributions between the control and rugby groups, as well as the forwards and backs sub-groups, for (A) *COL3A1* rs1800255, (B) *COL5A1* rs12722, (C) *COL6A1* rs35796750 and (D) *COL12A1* rs970547. Forwards sub-group, participants in playing positions 1-8, n=103. Backs sub-group, participants in playing positions 9-15, n=63.

8.3.3 Gene-gene Interactions and Rugby Union Playing Level and Position

No significant associations were identified when inferred pseudo-haplotypes, constructed from *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547, were investigated (Table 8.2).

Closer inspection of the four-gene inferred pseudo-haplotypes showed a tendency for certain combinations which included the same *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 alleles, such as T+T+G (A+T+T+G and G+T+T+G), to be over-represented in the backs sub-group. In addition there was at least a tendency for independent associations between these three collagen gene variants and the outside backs sub-group. Therefore inferred pseudo-haplotypes constructed from *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 were further investigated (Figure 8.2). The C+C+G inferred pseudo-haplotype was significantly (p=0.030, OR=2.3, 95% CI 1.0-5.4 and p=0.042, OR=2.3, 95% CI 0.8-6.1) under-represented in the control group (4.7%, n=8) when compared to the rugby group (10.1%, n=22) and backs sub-group (9.6%, n=9) (Figure 8.2). Furthermore, the T+T+G inferred pseudo-haplotype was significantly (p=0.029, OR=2.2, 95% CI 1.0-5.2) over-represented in the backs sub-group (13.2%, n=12) when compared to the control group (7.2%, n=13) (Figure 8.2).

Table 8.2. Frequency distributions of inferred pseudo-haplotypes, constructed from *COL3A1* rs1800255 G/A, *COL5A1* rs12722 T/C, *COL6A1* rs35796750 T/C and *COL12A1* rs970547 A/G, between the control and rugby groups, as well as the forwards and backs sub-groups.

Inferred pseudo-haplotype	Control Group (178)	Rugby Group (222)	Forwards Sub-group (129)	Backs Sub-group (93)	p-value[a]	p-value[b]	p-value[c]
G+C+C+G	3.8	9.5	8.0	10.6	0.118	0.164	0.108
G+C+C+A	8.0	12.0	13.2	10.4	0.267	0.380	0.524
G+C+T+G	5.4	5.4	3.6	8.1	0.565	0.872	0.090
G+C+T+A	17.0	8.5	8.9	6.7	0.089	0.304	0.065
G+T+C+G	4.6	4.1	4.6	5.3	0.982	0.980	0.121
G+T+C+A	11.5	10.1	4.1	12.8	0.885	0.959	0.311
G+T+T+G	4.5	5.8	1.7	7.1	0.965	0.987	0.508
G+T+T+A	18.4	17.3	24.1	16.2	0.344	0.811	0.210
A+C+C+G	2.6	0.0	0.0	0.0	nd	nd	nd
A+C+C+A	2.2	0.0	3.2	0.0	0.318	0.972	0.203
A+C+T+G	0.8	1.8	6.2	0.0	0.555	0.114	0.349
A+C+T+A	2.5	9.2	6.0	8.4	0.128	0.965	0.961
A+T+C+G	0.0	0.0	0.0	0.0	nd	nd	nd
A+T+C+A	8.7	6.8	10.1	3.1	0.112	0.490	0.103
A+T+T+G	3.7	4.1	3.0	6.5	0.905	0.311	0.311
A+T+T+A	6.4	4.3	2.0	4.7	0.786	0.078	0.133

Values are expressed as percentages. nd, not determined.

[a] control group vs. rugby group

[b] control group vs. forwards sub-group

[c] control vs. backs sub-group

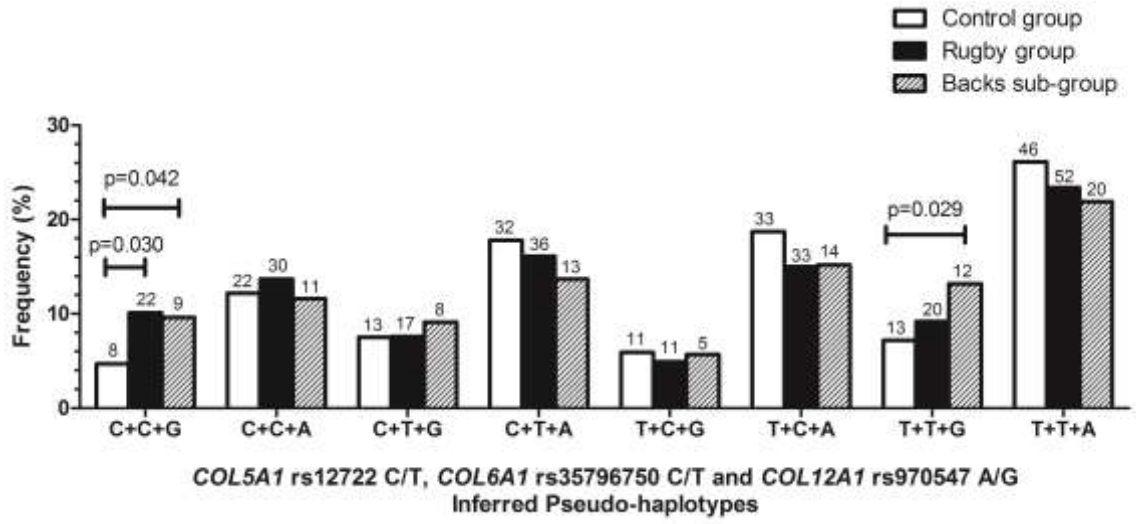


Figure 8.2. Frequency distributions of inferred pseudo-haplotypes, constructed from *COL5A1* rs12722 T/C, *COL6A1* rs35796750 T/C and *COL12A1* rs970547 A/G, between the control and rugby groups, as well as the backs sub-group. Numbers of participants are indicated above each column.

8.4 DISCUSSION

The first novel main finding of this study is that the *COL3A1* rs1800255 variant is independently associated with rugby union playing position. Specifically, the *COL3A1* rs1800255 GG genotype was significantly under-represented in the professional rugby playing forwards sub-group (playing positions 1-8 characterized by intense and strenuous activities [167]) when compared to control participants that had played at a recreational level or had never played rugby before. The *COL3A1* rs1800255 variant was selected since the G to A transition at this locus is proposed to result in an alanine to threonine change at position 698 of the COL3A1 peptide, which is proposed to affect the tensile strength of type III collagen fibres [94]. Interestingly, previous studies in this thesis have shown that the rs1800255 A allele, as part of a pseudo-haplotype, was associated with reduced risk of chronic Achilles tendinopathy (Chapter 4), while the G allele was associated with increased endurance swimming and running performance (Chapter 6). Endurance and power/sprint phenotypes are often associated with the opposite alleles of the same gene variant [26], and this may explain the differences described above.

Furthermore, the GA genotype was significantly over-represented in the professional rugby playing forwards sub-group when compared to the control group. It is possible that the heterozygous genotype was identified because the AA genotype is rare. Although rs1800255 is non-synonymous, these results do not exclude the possibility that this variant is tightly linked to an actual casual variant(s) within the *COL3A1* gene or a neighbouring gene. Future studies should investigate additional variants within this chromosomal region. No independent associations were identified for

COL5A1 rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 and rugby union playing position.

Independent investigation of *COL12A1* rs970547 in the smaller outside backs sub-group (playing position 11, 14 and 15) revealed that the GG genotype was significantly over-represented in the outside backs sub-group when compared to the control group, forwards sub-group and all other playing positions, while the *COL5A1* rs12722 TT and *COL6A1* rs35796750 TT genotypes showed similar trends. Caution needs to be taken when interpreting these results due to the small sample size of the outside backs sub-group. Further research is required to repeat these investigations in a larger cohort of professional outside back rugby union players.

The second novel finding of this study was that *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 interact to modulate rugby union playing position. The C+C+G inferred pseudo-haplotype was significantly over-represented in the rugby group and backs sub-group when compared to the control group. In addition, the T+T+G inferred pseudo-haplotype was also over-represented in the backs sub-group when compared to the control group. In addition, although this analysis was not included in the results section of this chapter, when the smaller outside backs sub-group (playing positions 11, 14 and 15) was analysed the expected C+C+A inferred pseudo-haplotype was significantly under-represented in the outside backs sub-group when compared to the control group, while the complementary T+T+G inferred pseudo-haplotype showed a tendency to be over-represented in the outside backs sub-group when compared to the control group. Further research is required to repeat these investigations in a larger cohort of

outside back rugby union professional players. In addition, no independent variant analyses, or gene-gene interactions, investigated in this study were associated with playing professional rugby union (playing level; control group vs. rugby group).

The main limitation of this study is the small sample sizes after dividing the rugby group into sub-groups such as backs or outside backs. This study has identified potential markers for rugby union playing position, however a larger cohort of professional players in required to accommodate larger sub-groups for these analyses. A second limitation to this study is the use of recreational rugby union players in the control group. These participants were added to increase the size of the control group to add statistical power, however this results in a less homogeneous control group which may have affected the ability of the study to determine small effects of the gene variants investigated. Additional studies should include a more homogeneous control group of participants with no history of ever playing rugby union. Finally, this study included only male participants. Despite the fact that rugby union is played predominantly by males, females also participate in the sport. Therefore future studies should investigate these gene variants in female cohorts to determine if the associations identified in this study may be gender-specific, as shown when some of these variants were investigated in other exercise-associated phenotypes [137;153;154] (Chapter 3).

In conclusion, this study identified *COL3A1* rs1800255 as a novel marker of rugby union playing position. Specifically, the *COL3A1* rs1800255 GG genotype was significantly under-represented in the professional rugby playing forwards sub-group (playing positions 1-8) when compared to control participants that had never played

rugby or had played at a recreational level. Furthermore, the GA genotype was significantly over-represented in the forwards sub-group when compared to the control group. Additional variants within this gene, and neighbouring genes, should be explored as additional genetic markers. Further studies require a larger rugby cohort to investigate specific positional sub-groups or individual rugby union playing positions. A summary of the results of this chapter are presented in table 8.4.

Table 8.4. A continuing summary of the results of the chapters in this thesis. The results of chapter 8, and all preceding chapters, are shown.

Investigated Phenotype	Associations	Investigated Gene Variants			
		COL3A1 **rs1800255 (G/A)**	***COL5A1*** **rs12722 (C/T)**	***COL6A1*** **rs35796750 (C/T)**	***COL12A1*** **rs970547 (A/G)**
ACL	**Genotype**	not associated	CC ♀ prev pub[1]	not associated	AA ♀ prev pub[2]
	Haplotype	nd	T ♀[3]	nd	A ♀[3]
Achilles tendinopathy	**Genotype**	not associated prev pub[4]	CC prev pub[5]	not associated	not associated prev pub[6]
	Haplotype	A	C	C	nd
ROM	**Genotype**	not associated	CC prev pub[7]	not associated	not associated
	Haplotype	not associated	C	T	not associated
Endurance Swim Performance[9]	**Genotype**	not associated	not associated	not associated	not associated
	Haplotype	G	C	T	not associated
Endurance Cycle Performance[9]	**Genotype**	not associated	not associated	TT	not associated
	Haplotype	not associated	not associated	not associated	not associated
Endurance Running Performance[9]	**Genotype**	not associated	TT prev pub[8]	not associated	not associated
	Haplotype	G	T	C	A
EAMC[9]	**Genotype**	not associated	CC	not associated	not associated
	Haplotype	not associated	C	T	not associated
Rugby Union[9]	**Genotype**	GG[10]	not associated	not associated	not associated
	Haplotype	not associated	T[11]	T[11]	G[11]

Green shading indicates that the allele or genotype is associated with a reduced risk of injury, except for ROM, endurance performance and rugby union where it represents an increased ROM or endurance performance and an over-representation in a position sub-group. Red shading indicates that the allele or genotype is associated with an increased risk of injury, except rugby union where it represents an under-representation in a position sub-group. Grey shading indicates no independent association. Prev pub, previously published. ♀, associations identified only in females. ACL, anterior cruciate ligament; ROM, range of motion; EAMC, exercise-associated muscle cramps; nd, not determined.

[1] independent association previously reported by Posthumus et al. [153].
[2] independent association previously reported by Posthumus et al. [154], but not independently associated in this thesis.
[3] T+A pseudo-haplotype also associated with increased risk of female ACL ruptures when a Polish cohort was analysed [137].
[4] no independent association previously reported by Saunders [170]
[5] independent association previously reported by Mokone et al. [131] and repeated by September et al. [179].
[6] no independent association previously reported by September et al. [181]
[7] independent association previously reported by Brown et al. [28;29].
[8] independent association previously reported by Posthumus et al. [151] and repeated by Abrahams et al. [1].
[9] only male participants were included in this study.
[10] Under-represented in the Rugby Union forwards sub-group in this thesis.
[11] Over-represented in the Rugby Union backs and outside backs sub-groups in this thesis.

CHAPTER 9

PRELIMINARY INVESTIGATION OF THE FUNCTIONAL EFFECTS OF *COL6A1* rs35796750

9.1 INTRODUCTION

As mentioned in chapter 2 of this thesis, the function of type VI collagen remains largely unknown. Although it is believed to play a role at the basement membrane, this has not been investigated [201]. The protein is expressed ubiquitously and is present in all connective tissue that contains types I and III collagen [106]. Furthermore it has been shown to interact with type V collagen in the fibril [190], and shows co-localisation with type III collagen in rat lung tissue [5]. As previously discussed a murine *Col6a1* knockout model for myopathy revealed that the average weekly distance run by *Col6a1 -/-* mice was consistently less than the wildtype mice [21].

The TT genotype of *COL6A1* rs35796750 (C/T) is associated with increased risk of a number of multifactorial conditions, such as ossification of the posterior longitudinal ligament (OPLL) [191] and diffuse idiopathic skeletal hyperostosis (DISH) [197] in independent Japanese populations. In this thesis the T and C alleles, as part of a haplotype with *COL5A1* rs12722, were significantly associated with reduced and increased risk of history of exercise-associated muscle cramps respectively (Chapter

7). In addition, the C allele was significantly over-represented in control participants, as part of a haplotype with the *COL3A1* rs1800255 and *COL5A1* rs12722 variants, when compared to participants with clinically diagnosed Achilles tendinopathy (Chapter 4).

As hypothesised, the *COL6A1* rs35796750 TT genotype was independently associated with increased endurance cycling performance in male ironman triathletes (Chapter 6). Consistent with this finding, the T allele, as part of a haplotype with *COL3A1* rs1800255 and *COL5A1* rs12722, was associated with increased endurance swimming performance during the ironman triathlon (Chapter 6). Furthermore, the T allele, as part of a haplotype with *COL5A1* rs12722 and *COL12A1* rs970547, was significantly over-represented in male rugby union outside backs when compared to controls (Chapter 8). The C allele, on the other hand, was associated, as part of a haplotype with *COL3A1* rs1800255, *COL5A1* rs12722 and *COL12A1* rs970547, with increased endurance running performance.

The *COL6A1* rs35796750 gene variant occurs near the branch site of intron 32 where the cytosine to thymine transition is proposed to cause aberrant splicing of the COL6A1 mRNA [191]. Similarly to the previously published functional studies on *COL5A1* rs12722 [107], it is important to determine the functional and biological mechanisms, of how *COL6A1* rs3579750 may modulate exercise-related and other multifactorial phenotypes. The aim of this study was therefore to perform a preliminary investigation of the functional effects of *COL6A1* rs35796750. Specifically, to determine if *COL6A1* gene expression was different between the C allele and T allele of the rs35796750 varaint.

9.2 MATERIALS AND METHODS

9.2.1 Bioinformatics Analysis of COL6A1 rs35796750

A 601 bp sequence of the *COL6A1* gene, containing rs35796750, was interrogated using a number of online bioinformatics tools, including TESS [173;187], ESEfinder [36] and JASPAR [30]. These databases identify conserved transcription factor [36;173;187] and splice factor [30] DNA-binding sites within the target sequence.

9.2.2 Participants and Tissue Samples

A total of six participants, with known *COL6A1* rs35796750 (C/T) genotypes, were recruited for this preliminary study. Two participants were recruited with each genotype (TT, TC and CC). Primary skin fibroblasts were isolated and cultured from skin biopsy samples taken from these participants using methods described by Baumgarten [12]. Briefly, the dermis and epidermis were removed from the biopsy sample and washed with sterile phosphate buffered saline (PBS). This tissue was then finely sliced into approximately 1 mm by 1 mm pieces while covered in Dulbecco's modified eagle medium (DMEM) (Highveld Biological, South Africa), 10% fetal calf serum (FCS) (Highveld Biological, South Africa), 200 units/ml penicillin and 100 µg/ml streptomycin (complete medium). The sliced tissue fragments were then placed in 35 mm culture dishes and covered with sterile coverslips, thereby keeping the fragments in place and allowing them to attach to the dish surface. Furthermore, 2.5 ml of complete medium were then added to each culture dish. The dishes were

the placed in a 37 ^{o}C humidified incubator (95% air, 5% CO_2, 65% humidity). The medium was changed twice weekly until enough cells had grown out of the explants. Fibroblasts growing out form the explant can be seen in figure 9.1.

Once the cells had grown out of the explants the coverslips were transferred, cell side up, into new 35 mm dishes. After approximately 2-3 weeks and the cells grown from the explants covered 70-90% of the culture area the medium was removed. The cells were then washed twice with PBS, then trypsinised using a 0.25% trypsin/EDTA solution and seeded in 50 ml culture flasks (passage one). Testing for mycoplasma was performed between passages on cells grown in antibiotic-free medium containing 10% FCS. Subcultures involved expansion of the fibroblast cultures into several flasks in complete medium. Stocks of fibroblasts from passage two and three were cryogenically preserved using 10% DMSO (Sigma-Aldrich, Munich, Germany) and stored in liquid nitrogen for further usage. The primary fibroblast cells obtained from early passages (Passages 3-6) were seeded at 2 X 10^5 cells per ml in two 6 cm dishes per samples in 4 ml of complete DMEM. These cells were then incubated in a 37 ^{o}C humidified incubator (95% air, 5% CO_2, 65% humidity) overnight or until an 80% confluency was obtained.

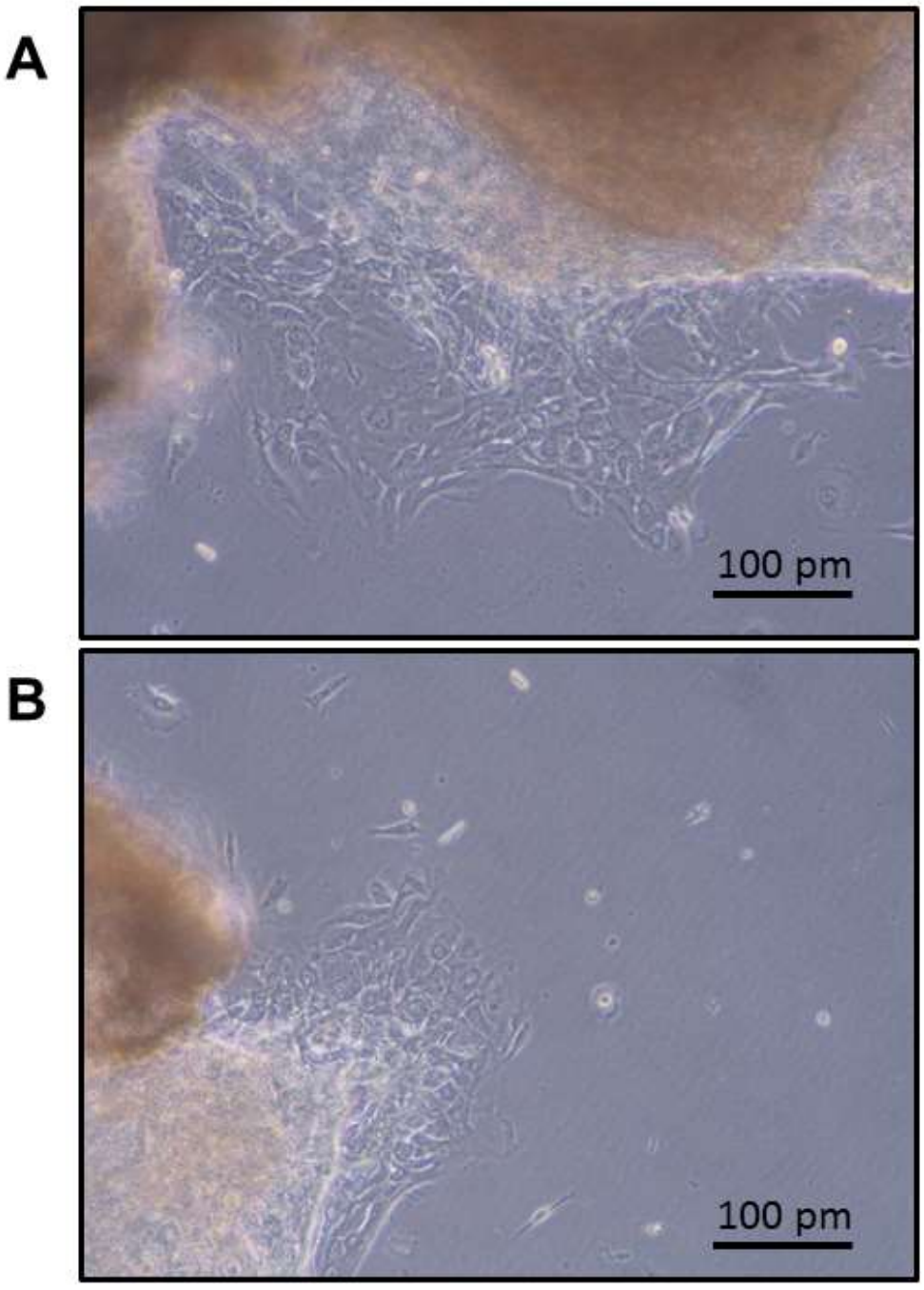


Figure 9.1. Panels A and B show fibroblasts growing from the explant after 1 week under high magnification.

9.2.3 Total RNA isolation and cDNA synthesis

Total RNA was extracted from primary fibroblasts using a High Pure RNA Isolation Kit (Roche Diagnostics, Roche Applied Science, Mannheim, Germany), according to the manufacturer's specifications with minor modifications to increase the concentration of the RNA yield. Briefly; primary fibroblasts were resuspended in 100µl PBS. The samples were then vortexed for 15s following the addition of 400µl Lysis/-Binding Buffer. Each sample was then transferred to a High Pure Filter Tube, inserted into a Collection Tube, and centrifuged at 8000 x g for 15s. After centrifugation the flowthrough liquid was discarded. A total of 90µl DNase Incubation Buffer and 10µl DNase I was then added to each sample and they were left to incubate for 15 min at room temperature. Following this incubation step, 500µl Wash Buffer I was added to each sample and they were centrifuged for 15s at 8000 x g. Again the flowthrough liquid was discarded. 500µl of Wash Buffer II was then added and the samples were centrifuged at 8000 x g for 15s. After discarding the flowthrough liquid a final 200µl of Wash Buffer II was added to each sample. The samples were then centrifuged at 13 000 x g for 2min. Following this final wash step the High Pure Filter Tubes were transferred to sterile 1.5ml microcentrifuge tubes. To elute the RNA, 30µl of Elution Buffer was added to each sample and they were centrifuged at 8000 x g for 1min. The microcentrifuge tube then contained the RNA and was stored at -80^{o}C for later analysis.

Conventional reverse transcription was performed using the ImProm-II Reverse Transcription System (Promega Corporation, Madison, Wisconsin, USA), according to the manufacturer's specifications. Briefly, 1µg of RNA per sample was added to

1µl of random and oligo(dT) primers in a final reaction volume of 5µl. These reaction tubes were then placed on a 70^{o}C heating block for 5min before being chilled in ice-water for at least 5min. While these samples were chilling a 15µl reverse transcription reaction mix was made containing; 6.1µl nuclease-free water, 4.0µl ImProm-II 5X Reaction Buffer, 2.4µl of $MgCl_2$ (final concentration 3mM), 1µl of dNTP Mix (final concentration 0.5mM each dNTP), 0.5µl Recombinant RNasin Ribonuclease Inhibitor and 1.0µl ImProm-II Reverse Transcriptase. These 15µl aliquots were then added to the 5µl sample specific reactions made previously. The 20µl reaction tubes were then placed on a heating block at 25^{o}C for 5min, incubated at 42^{o}C for 1 hour and then heated at 70^{o}C for 15min to inactivate the reverse transcriptase enzyme. The cDNA samples were then stored at -20^{o}C until analysed.

9.2.4 Quantitative Real-Time PCR

Quantitative Real-time PCR (qRT-PCR), to determine the relative content of COL6A1 and COL1A1 mRNA, was performed using the 7900 HT Fast Real-Time PCR System and SDS Software version 2.3 (Applied Biosystems, Foster City, CA, USA). Commercially available fluorescence-based custom Taqman® array assays were used (Appendix D) (Applied Biosystems, Foster City, CA, USA). Each sample was analysed in triplicate, in two separate experiments, using the following PCR conditions. A 10 min heat activation step (95ºC) followed by 40 cycles of 15 sec at 92ºC and 1 min at 60ºC. Fluorescence data, indicative of the amount of PCR product, was captured at each cycle. The relative COL6A1 and COL1A1 mRNA concentrations were then calculated based on the cycle number that the threshold quantity of PCR product is obtained (C_t). C_t values for COL6A1 and COL1A1 were

normalised to values of an internal housekeeping gene, β-Actin, and expressed relative to this control [116]. The original and normalised C_t values are presented in Appendix C. Finally, at the end of the PCR, a melting curve was performed to confirm single melting peaks, indicative of a single PCR product, for each reaction.

9.2.5 Statistical Analysis

The same statistical tests were performed as described in chapter 3 section 3.2.4. Significance was accepted when $p<0.05$.

9.3 RESULTS

9.3.1 Identified DNA-binding sits at rs35796750

The results of the interrogation of the gene region containing *COL6A1* rs35796750, using online bioinformatics tools [30;36;173;187], are shown in figure 9.2. Specifically, the DNA-binding domains for splice factor SRp55 (sequence TGCGCC) and the E2F transcription factor family (sequence GCGCC) were identified in the presence of *COL6A1* rs35796750 C allele, while both sites are abolished in the presence of the T allele, respectively (Figure 9.2).

Figure 9.2. DNA-binding domains identified at the *COL6A1* rs35796750 locus. The splice factor SRp55 (sequence TGCGCC) and the E2F transcription factor family (sequence GCGCC) were identified in the presence of the *COL6A1* rs35796750 C allele, while both sites are abolished in the presence of the T allele, respectively.
- TGCGCC, SRp55 splice factor binding site.
- GCGCC, E2F transcription factor family binding site.

9.3.2 Participant Characteristics

The physical characteristics of the individual participants, as well as by *COL6A1* rs35796750 genotype group, are presented in table 9.1.

Table 9.1. General characteristics for the individual participants, as well as by *COL6A1* rs35796750 genotype group, recruited in this study.

Participants	Age (yrs)	Weight (kg)	Height (cm)	Gender	Genotype Group [a]	Age (yrs)	Weight (kg)	Height (cm)
TT1	47	73	173	M	**TT**	36.5 ± 14.8	74.6 ± 2.1	178.0 ± 7.1
TT2	26	76	183	M				
TC1	22	65	174	F	**TC**	26.0 ± 5.7	72.5 ± 10.6	168.5 ± 7.8
TC2	30	80	163	F				
CC1	24	96	179	M	**CC**	25.0 ± 1.4	91.0 ± 7.1	176.5 ± 3.5
CC2	26	86	174	F				

Where appropriate, values are expressed with mean ± standard deviation. M, male. F, female.

[a] *COL6A1* rs35796750 genotype groups

9.3.3 Collagen Gene Expression Levels

The individual gene expression, relative to participant CC1, for *COL6A1* and *COL1A1* are presented in figure 9.3. Individual differences between the relative expression levels for *COL6A1* and *COL1A1* were identified (Figure 9.3). Specifically, participant TT2 showed significantly (p=0.001) higher *COL6A1* expression when compared to *COL1A1* expression, while a similar tendency (p=0.088) for higher *COL6A1* expression was also identified for participant TT1 (Figure 9.3). No other

significant differences between the individual participant expression levels of *COL6A1* and *COL1A1* were identified.

Inter-individual differences in the relative expression of *COL6A1* and *COL1A1* were also identified. Participant TC1 showed a significant ($p=0.016$) 2.7 ± 1.5 fold increase in *COL6A1* expression when compared to participant CC1, while a tendency ($p=0.052$) was observed for participant TT1 to have a higher *COL6A1* expression (2.5 ± 1.2 fold change) when compared to participant CC1 (Figure 9.3). In addition, participant TC1 also showed a significant ($p<0.001$) 2.6 ± 1.1 fold increase in *COL1A1* expression when compared to participant CC1 (Figure 9.3B). No other significant differences were identified for either *COL6A1* or *CO1A1* gene expression relative to participant CC1.

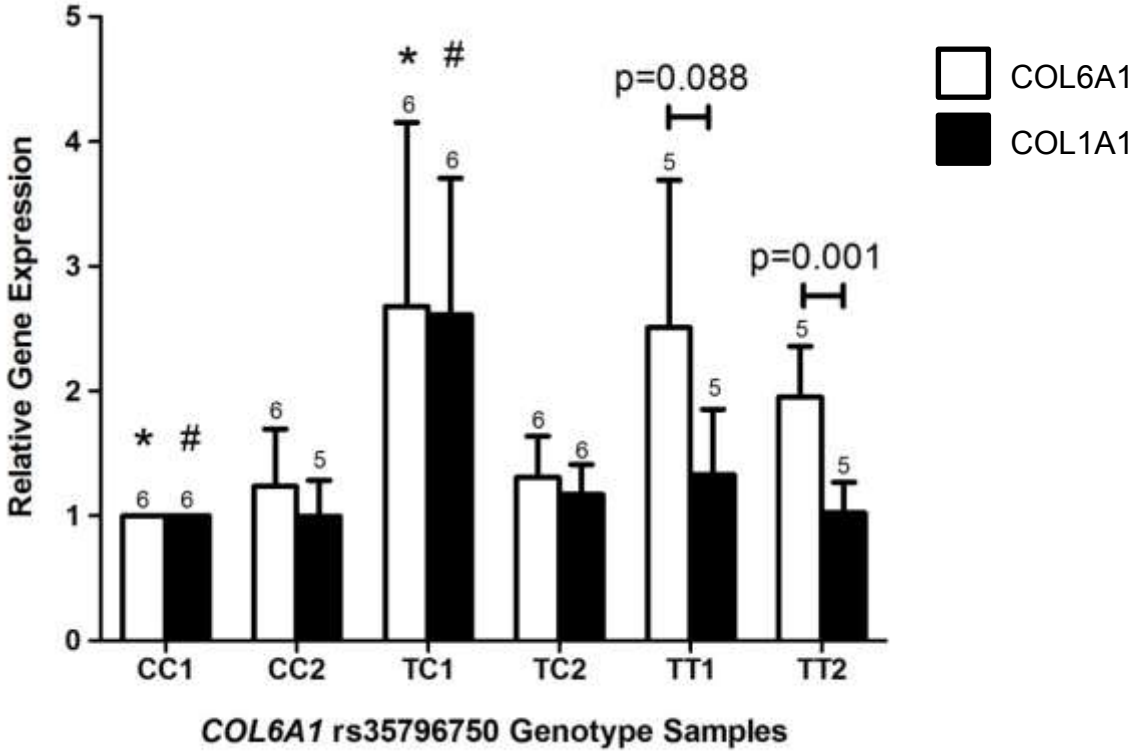


Figure 9.3. The individual participant expression levels of *COL6A1* (white bars) and *COL1A1* (black bars) relative to participant CC1. The number of experimental replicates is shown above each column. p-values are presented above relevant columns. The star (*) indicates a significant (p=0.016) difference in the relative expression of *COL6A1* between participants TC1 and CC1. The hash (#) indicates a significant (p<0.001) difference in the relative expression of *COL1A1* between participants TC1 and CC1.

The expression levels of *COL6A1* and *COL1A1*, after grouping participants according to their *COL6A1* rs35796750 genotype, relative to the CC genotype group are presented in figure 9.4. When the relative expression levels of *COL6A1* and *COL1A1* were compared within the same *COL6A1* rs35796750 genotype group a significant difference was identified for the TT genotype group. Specifically, the relative expression level of *COL6A1* (1.89 ± 0.54) was significantly (p=0.003) higher when compared to that of *COL1A1* (1.17 ± 0.41) (Figure 9.4). The relative expression levels of *COL6A1* and *COL1A1* were not different within the CC (p=0.304) and TC (p=0.830) genotype groups (Figure 9.4).

Significant differences in *COL6A1* expression were also identified between the three genotype groups. Specifically, the *COL6A1* rs35796750 TT (1.89 ± 0.54) and TC (1.76 ± 0.84) genotype groups had significantly (p=0.005 and p=0.013) higher *COL6A1* expression levels when compared to the CC (1.02 ± 0.21) genotype group, respectively (Figure 9.4). Similarly, significant differences in *COL1A1* expression were also identified between the three *COL6A1* rs35796750 genotype groups. Specifically, the *COL6A1* rs35796750 TC (1.90 ± 1.10) genotype group had significantly (p=0.013) higher *COL1A1* expression levels when compared to the CC (1.01 ± 0.16) genotype group (Figure 9.4), while the TT (1.17 ± 0.41) genotype showed a tendency (p=0.053) to be higher (Figure 9.4).

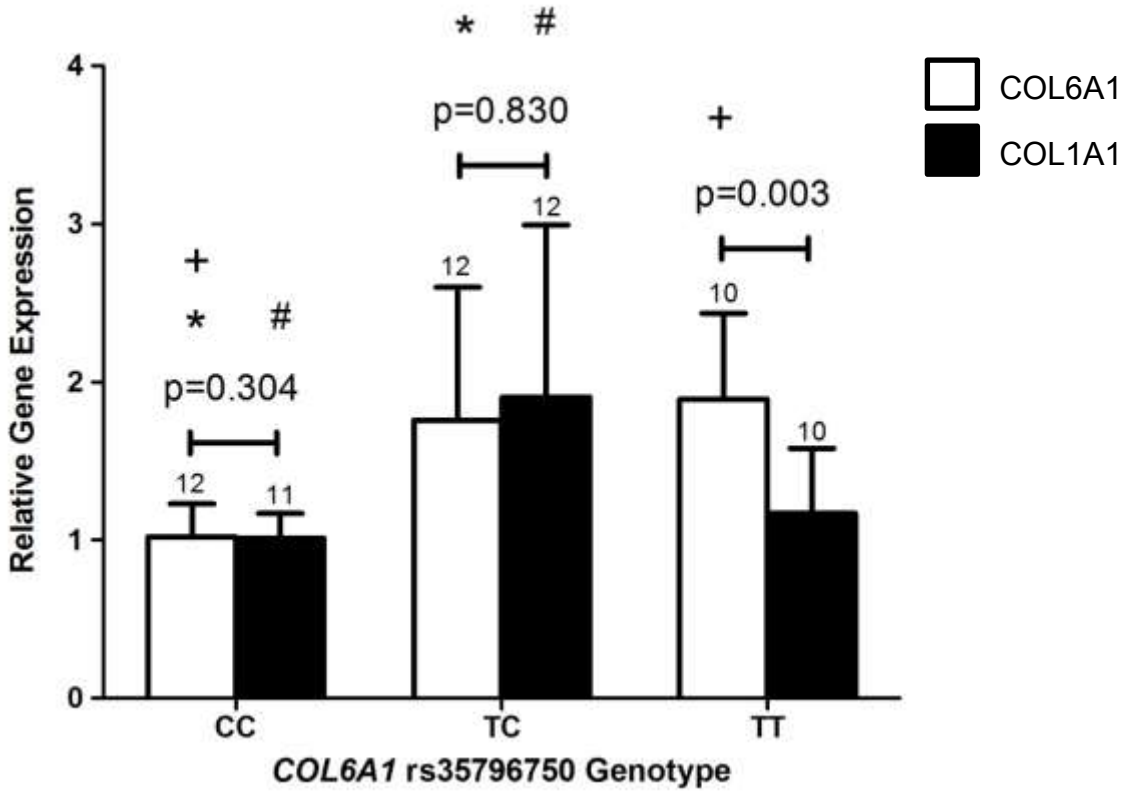


Figure 9.4. The *COL6A1* rs35796750 genotype group expression levels of *COL6A1* (white bars) and *COL1A1* (black bars) relative to participants with the *COL6A1* rs35796750 CC genotype. The number of experimental replicates is shown above each column. p-values are presented above relevant columns. The star (*) indicates a significant (p=0.013) difference in the relative expression of *COL6A1* between participants in the CC and TC genotype groups. The plus (+) indicates a significant (p=0.005) difference in the relative expression of *COL6A1* between participants in the CC and TT genotype groups. The hash (#) indicates a significant (p=0.013) difference in the relative expression of *COL1A1* between participants in the CC and TC genotype groups.

9.4 DISCUSSION

The novel preliminary results of this study showed that participants with the *COL6A1* rs35796750 TT and TC genotypes had significantly higher *COL6A1* gene expression when compared to participants with the CC genotype, suggesting that individuals with at least one *COL6A1* T allele have a higher type VI collagen gene expression. However analysis of *COL1A1* gene expression between the *COL6A1* genotype groups showed a similar *COL1A1*:*COL6A1* expression ratio in individuals with the *COL6A1* rs35796750 TC and CC genotypes, with a higher *COL1A1*:*COL6A1* ratio in the participants with a *COL6A1* TT genotype, suggesting that relative to type I collagen gene expression, type VI collagen might be higher in individuals with a *COL6A1* TT genotype.

Functional variants, such as the Sp1-binding site within intron 1 [81;126], within the *COL1A1* gene alter its expression at an individual level. This could not be accounted for in this study. Despite this, it is interesting to note that the level of *COL6A1* gene expression was higher when compared to *COL1A1* gene expression in both the participants with a *COL6A1* TT genotype. Although very speculative, this could suggest that the observed increased *COL6A1* gene expression is due to the rs35796750 TT genotype and not due to altered *COL1A1* gene expression. However, future studies with larger samples sizes should genotype participants for all known functional variants within the *COL1A1* gene to eliminate altered *COL1A1* gene expression as a possible confounder.

Based on the results of this study two models can be proposed to explain the mechanism through which the rs35796750 variant may result in altered *COL6A1* gene expression. The first model assumes the altered binding affinity of the SRp55 splice factor to *COL6A1* at the polymorphic intronic DNA-binding domain containing the rs35796750 variant. The regulation of splicing by SR-proteins and their kinases has been extensively reviewed by Zhou and Fu [216]. Briefly, an SR-protein binding to an exonic DNA-binding site is positive for that exons inclusion during splicing, however binding to an intronic DNA-binding site will result in exon exclusion during splicing. For example, the SRp55 splice factor binds to the target sequence TGCGCC within intron 32 of *COL6A1*, in the presence of the rs35796750 C allele, which would result in exclusion of exon 33 during splicing (Figure 9.5). The binding site is abolished in the presence of the rs35796750 T allele which would result in inclusion of exon 33 (Figure 9.5). This exon exclusion would result in a smaller COL6A1 transcript.

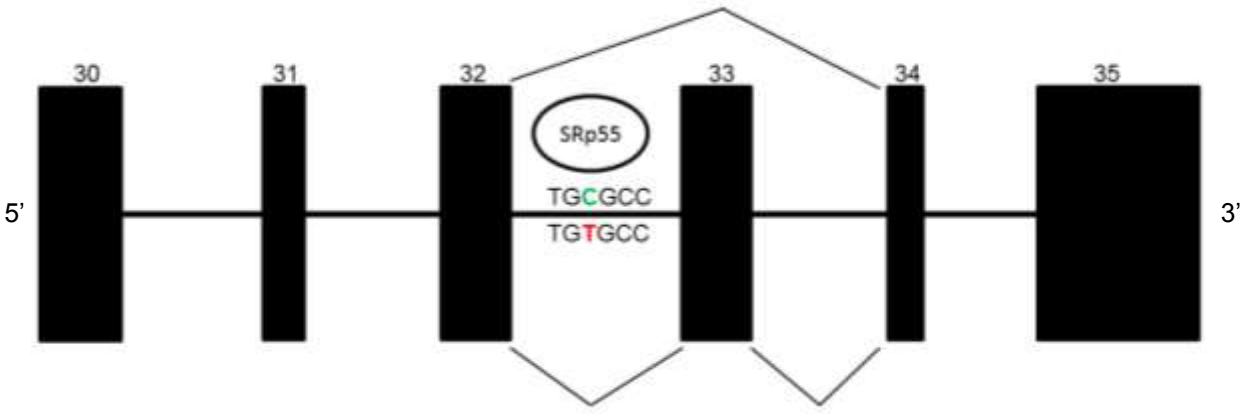


Figure 9.5. A schematic diagram showing alternative splicing of the *COL6A1* gene as a result of the SRp55 splice factor binding to the intronic region containing the rs35796750 variant. The last six exons (30-35) of the *COL6A1* gene are indicated as black rectangles while introns are represented by the black line between exons. The SRp55 splice factor binds to the intronic DNA-binding domain, containing the rs35796750 C allele (in green), resulting in exclusion of exon 33 during splicing. The SRp55 splice factor does not bind in the presence of the rs35796750 T allele (in red) resulting in inclusion of exon 33 during splicing.

Alternative α3(VI) collagen transcripts are known to vary in abundance, with no tissue-specification, in chick tissues [47]. Similar results for α3(VI) collagen transcripts of differing sizes have also been shown in a murine model [48]. It is therefore tempting to speculate that alternate splicing, as a result of rs35796750, may also result in varying abundance of the α1(VI) collagen transcript (Figure 9.6). Specifically, the binding of SRp55 to the DNA-binding domain in the presence of the rs35796750 C allele may result in exon exclusion and thereby lead to reduced α1(VI) mRNA abundance which may in turn lead to reduced *COL6A1* gene expression (Figure 9.6). The inclusion of *COL6A1* exon 33, as a result of the rs35796750 T allele and the abolishment of the SRp55 DNA-binding site, would then result in a larger α1(VI) collagen transcript which may lead to increased α1(VI) mRNA abundance which may in turn lead to increased *COL6A1* gene expression (Figure 9.6).

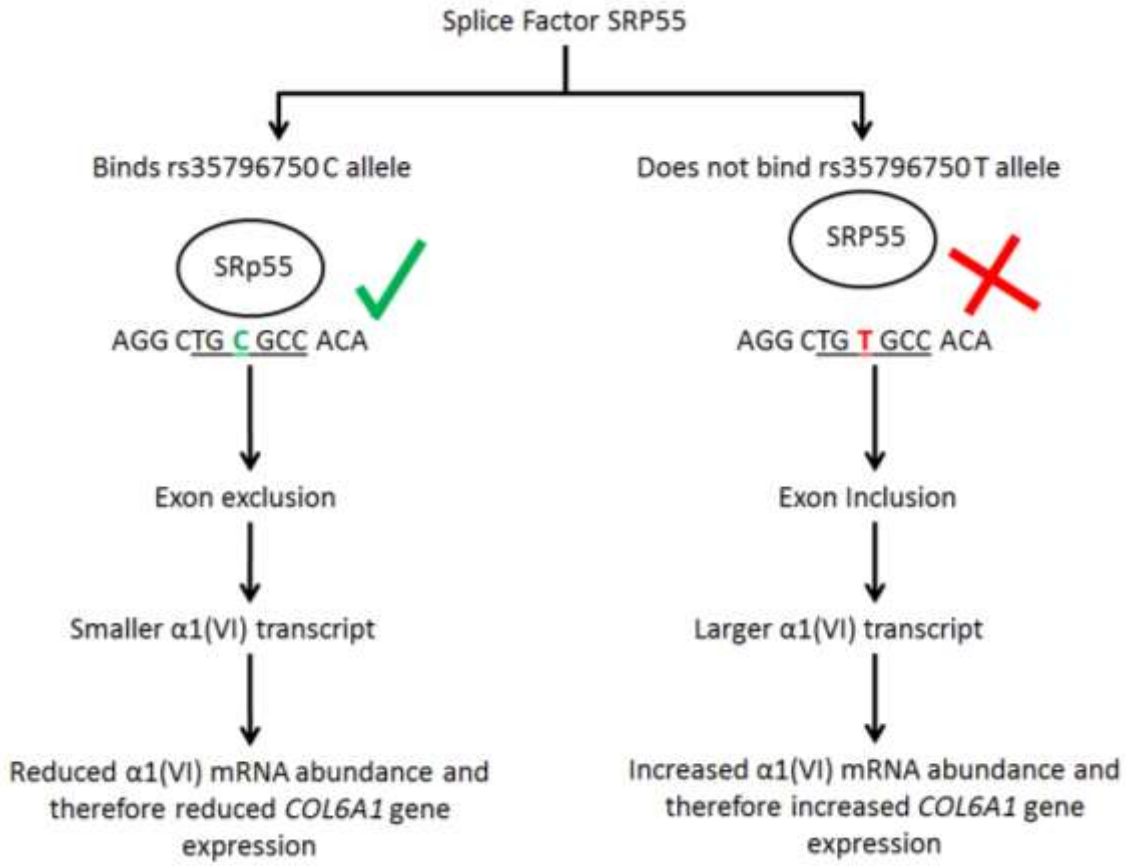


Figure 9.6. A schematic diagram of the proposed interaction between the intronic *COL6A1* gene DNA-binding site, containing rs35796750, and the SRp55 splice factor. In the left panel SRp55 binds to the DNA-binding site in the presence of the rs35796750 C allele (green) which may then lead to reduced α1(VI) mRNA abundance and corresponding *COL6A1* gene expression. The right panel shows that the binding site is abolished in the presence of the rs35796750 T allele (red) resulting in increased α1(VI) mRNA abundance and corresponding *COL6A1* gene expression. The DNA-binding site is underlined.

The second model that may explain the altered *COL6A1* gene expression as a result of rs35796750 assumes the altered binding affinity of an E2F transcription factor to *COL6A1* at the intronic DNA-binding domain containing the rs35796750 variant. The family of E2F transcription factors recognise and bind the target sequence GCGCC within intron 32 of the *COL6A1* gene in the presence of the rs35796750 C allele. This DNA-binding site sequence is abolished in the presence of the rs35796750 T allele (Figure 9.7). E2F transcription factors act as activators (E2Fs 1-3) which promote gene expression or as inhibitors (E2Fs 4-8) which act as transcriptional repressors [8]. Therefore is may be proposed that an inhibitory E2F transcription factor binds to the intronic DNA-binding site in the presence of the rs35796750 C allele resulting in *COL6A1* transcriptional repression thereby reducing *COL6A1* gene expression (Figure 9.7). The presence of the rs35796750 T allele abolishes the E2F binding site thereby preventing transcriptional repression and resulting in increased *COL6A1* gene expression (Figure 9.7).

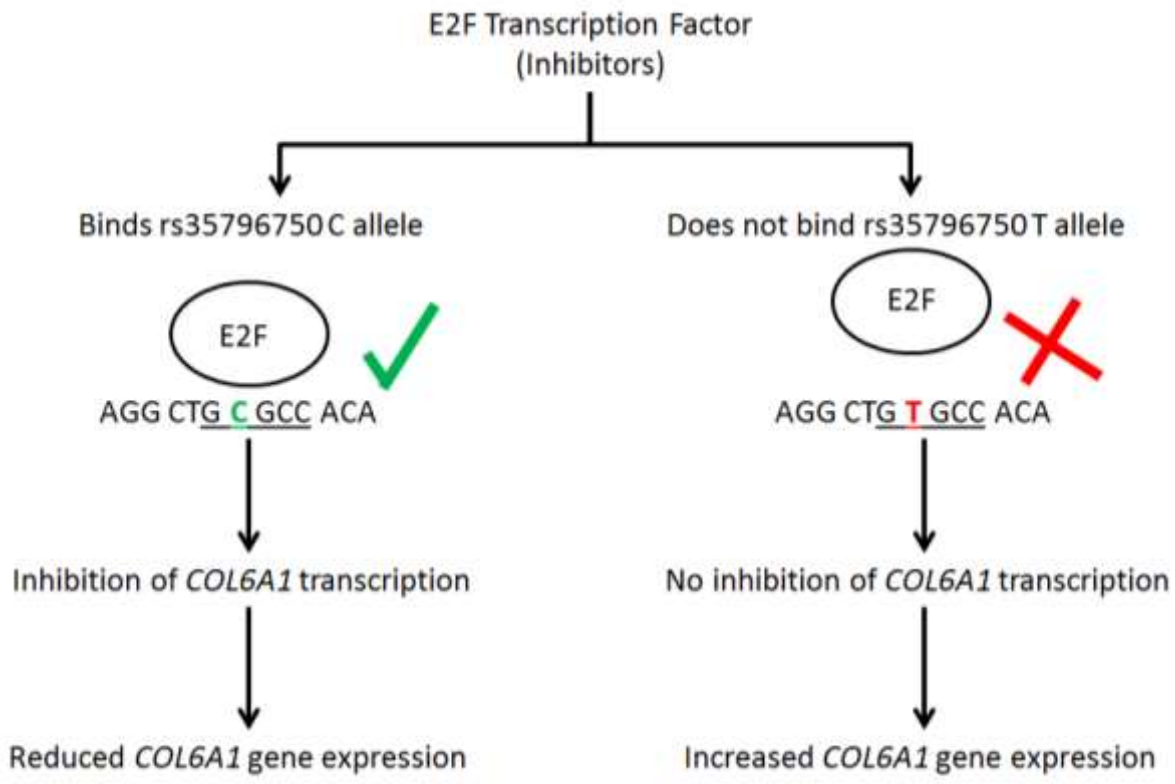


Figure 9.7. A schematic diagram of the proposed interaction between the intronic *COL6A1* gene DNA-binding site, containing rs35796750, and the E2F transcription factor. In the left panel an inhibitory E2F transcription factor binds to the DNA-binding site in the presence of the rs35796750 C allele (green) which may then lead to transcriptional repression and corresponding reduced *COL6A1* gene expression. The right panel shows that the binding site is abolished in the presence of the rs35796750 T allele (red) resulting in a lack of transcriptional repression and a corresponding increased *COL6A1* gene expression. The DNA-binding site is underlined.

It is therefore possible that a continuum of associated phenotypes may also exist for the range of genetic variation within the *COL6A1* gene, similar to that proposed for the *COL5A1* gene in chapter 2 section 2.2 [41]. Mutations within a single copy of the *COL6A1* gene will result in Bethlem myopathy and Ullrich congenital muscular dystrophy [109]. At this end of the continuum these severe disorders occur regardless of environmental exposure and independent of other non-genetic risk factors (Figure 9.8). At the opposite end of the continuum, functional common variants within *COL6A1*, such as rs35796750, contribute to more complex and less severe phenotypes, and do not result in disorders (Figure 9.8). These multifactorial phenotypes arise as a result of the interaction between genetic and non-genetic factors modifying physiological responses to environmental exposures [41].

A murine model has shown that *Col6a1 -/-* null mice have collagen fibrils with smaller diameters and greater density when compared to wildtype mice, accompanied by reduced tendon stiffness and maximal load [78]. Furthermore, the average weekly distance run by these *Col6a1 -/-* mice was consistently less than the wildtype mice [21], highlighting that these architectural changes may result in changes in biomechanics. Interestingly, analysis of *COL6A1* rs35796750 C/T in this study has shown that participants with the TT genotype had significantly higher *COL6A1* gene expression when compared to participants with the CC genotype. It may therefore be proposed that the increase in *COL6A1* gene expression, as a result of the rs35796750 TT genotype, may result in an increased collagen fibril diameter and reduced fibril density, which in turn may increase the stiffness and maximal load of the type VI collagen containing tissue including tendons, ligaments and muscle (Figure 9.8). In contrast, the CC genotype may reduce *COL6A1* gene expression

resulting in reduced collagen fibril diameter and a corresponding increase in fibril density (Figure 9.8).

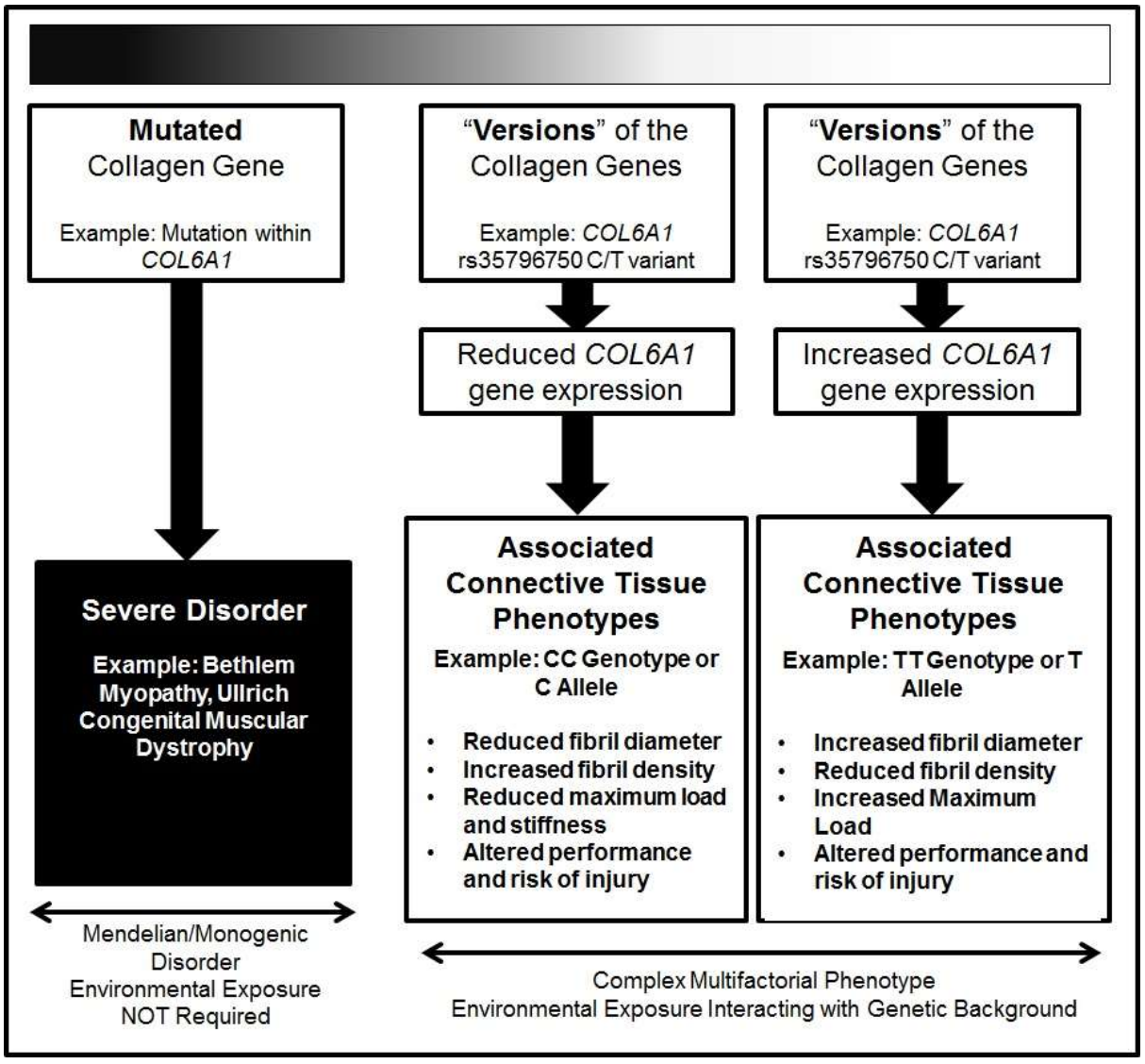


Figure 9.8. A proposed genetic continuum for the *COL6A1* gene, adapted from the general genetic continuum proposed by Collins and Posthumus [41]. The black shading represents the severe disorders, including Bethlem myopathy and Ullrich congenital muscular dystrophy [109], due to mutations in the *COL6A1* gene. At this end of the continuum a single mutation results in the disorder. Murine models have shown that *Col6a1* null mice have aberrant changes to collagen fibril size or diameter and fibril density [78]. The white shading represents the most beneficial "versions" of the *COL6A1* gene. At this end of the continuum variants within the gene collectively contribute to the aetiology of the phenotype. The *COL6A1* rs35796750 variant is shown to be functional in this study. Furthermore, this thesis has shown this functional variant to associate with a number of musculoskeletal soft tissue injuries and other exercise-related phenotypes.

This suggests that participants with the rs35796750 TT genotype or T allele should have increased athletic performance and reduced risk of musculoskeletal soft tissue injury when compared to participants with the CC genotype or C allele. Some of the studies within this thesis agree with this proposal (Chapters 6-8) and it may also explain previously described genetic associations in which the C allele of *COL6A1* rs35796750 was associated with increased risk of both OPLL [191] and DISH [197] in independent Japanese populations. However contrasting results have also been identified (Chapters 4 and 6).

These contrasting results may highlight tissue specific functions for this variant. Within phenotypes characterised by skeletal muscle involvement (endurance swimming and cycling, rugby union and EAMC) the associations identified fit the mechanism proposed above, the *COL6A1* rs35796750 T allele is associated with increased performance and reduced risk. Similarly, for OPLL and DISH which are both characterised by ligament involvement, the C allele was associated with increased in independent Japanese populations [191;197]. Interestingly, however, phenotypes characterised by tendon involvement (chronic Achilles tendinopathy and endurance running) show opposite associations in that the C allele is associated with increased performance and reduced risk of injury. Future studies are required to further investigate these findings. Furthermore, these contrasting results may be the result of the multi-factorial and polygenic nature of the phenotypes investigated in this thesis which will be discussed in more detail in the final concluding chapter of this thesis.

One of the main limitations to this study is the small sample used. Larger samples sizes should be used to repeat this investigation to determine that the findings of this study are accurate for the *COL6A1* rs35796750 genotype groups. Furthermore, this study was merely a preliminarily investigation of rs35796750 and further studies are required to determine exactly how this variant modulates *COL6A1* gene expression. Two possible mechanisms have been proposed above however these remain speculative until further studies are conducted.

This is the first study to show a functional effect on *COL6A1* expression as a result of the rs35796750 gene variant. Specifically, participants with the TT genotype had significantly higher *COL6A1* gene expression when compared to participants with the CC genotype. In addition, this study proposes a model to explain how the functional rs35796750 gene variant, and other *COL6A1* gene variants, may result in altered exercise-associated phenotypes.

CHAPTER 10

SUMMARY AND PERSPECTIVES

10.1 NOVEL FINDINGS OF THIS THESIS

Genetic variants within collagen genes have been previously associated with musculoskeletal soft tissue injuries and other exercise-related phenotypes (Chapter 2). Initially, the *COL5A1* rs12722 gene variant was associated with risk of Achilles tendinopathy [131;179] and ACL rupture in females [153]. Since mutations within the *COL5A1* gene cause classic type EDS [124], which is characterised by, among other characteristics, joint hypermobility and hyper-extensibility, additional studies investigated and identified associations between *COL5A1* gene variants and ROM [1;28;29;40], and later endurance performance [1;151] (Chapter 2 Section 2.3.2). Furthermore, it was proposed that the relative content of type V collagen in tendons, ligaments and other tissues alters the fibril diameters and packing density within these tissues, and may alter their mechanical properties, and therefore their susceptibility to injury and other exercise-related phenotypes [41]. These studies highlighted the collagen component as a factor in the aetiologies of these phenotypes. Similarly to *COL5A1*, mutations within the *COL3A1*, *COL6A1* and *COL12A1* genes cause severe connective tissue disorders [17;109;140;186;218]. In addition, the collagens for which these genes encode, types III, VI, and XII, are involved in similar biological processes as type V collagen [56;115;129;141;213].

Therefore, investigation of candidate variants in other collagen genes, such as *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547, was required to determine how they might modulate the aetiologies of these musculoskeletal soft tissue injuries and other exercise-related phenotypes. Furthermore, investigation of collagen gene variants in novel candidate exercise-related phenotypes was required to further determine the effects that these variants may have. In this concluding chapter, the novel findings, which address the aims of this thesis (Chapter 2 Section 2.7) will be summarised and the perspective and future direction will be discussed.

10.1.1 Primary Aim: To investigate the COL3A1 rs1800255 and COL6A1 rs35796750 gene variants as risk factors for ACL ruptures and chronic Achilles tendinopathy

The *COL5A1* rs12722 gene variant has been previously associated with risk of anterior cruciate ligament (ACL) ruptures in females [153] and with chronic Achilles tendinopathy [131]. The *COL12A1* rs970547 gene variant has also previously been associated with risk of ACL ruptures in females [154], but not with chronic Achilles tendinopathy [181]. Furthermore, the *COL3A1* rs1800255 gene variant was not associated with risk of chronic Achilles tendinopathy [170]. The *COL3A1* and *COL6A1* genes have not been previously investigated for risk of ACL ruptures, and the *COL6A1* gene has not been investigated for risk of chronic Achilles tendinopathy. Therefore, where appropriate, this thesis investigated the *COL3A1* rs1800255 and *COL6A1* rs35796750 gene variants as risk factors for these musculoskeletal soft tissue injuries.

The novel findings of this thesis were that the *COL3A1* rs1800255 and *COL6A1* rs35796750 gene variants were not independently associated with risk of ACL ruptures. In addition, no significant gene-gene interactions between these variants and risk of ACL rupture were identified. These results therefore suggest that *COL3A1* rs1800255 and *COL6A1* rs35796750 are not risk factors for ACL ruptures.

The novel findings of this thesis were that the *COL6A1* rs35796750 variant was not independently associated with risk of chronic Achilles tendinopathy in a South African or Australian cohort in this thesis. However, a novel gene-gene interaction between *COL6A1* rs35796750 and the previously investigated *COL3A1* rs1800255 variant [170] was shown to modulate the risk of Achilles tendinopathy.

10.1.2 Secondary Aim: To investigate the COL3A1, COL6A1, and COL12A1 genes for associations with range of motion and endurance performance

Of the collagen genes investigated in this thesis, only the *COL5A1* rs12722 variant was previously associated with endurance performance [151] and range of motion (ROM) [28;29]. Therefore, this thesis investigated the *COL3A1*, *COL6A1*, and *COL12A1* genes for associations with these exercise-related phenotypes.

The novel findings of this thesis were that the *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547 variants were not independently associated with specific upper and lower body joint range of motion (ROM) measurements. Despite a lack of independent associations, the *COL3A1* rs1800255 and *COL6A1* rs35796750 were shown to interact again to modulate individual sit-and-reach ROM

variation. These results implicate these genes in the aetiology of ROM for the first time.

A further novel finding of this thesis was the identification of the *COL6A1* rs35796750 gene variant as a novel performance marker for the bike component of the South African Ironman triathlon. The *COL6A1* rs35796750 TT genotype was associated with increased performance during the bike component of the South African Ironman triathlon. No independent associations were identified between *COL3A1* rs1800255 or *COL12A1* rs970547 and time to complete any of the components of the Ironman triathlon. However, *COL3A1* rs1800255 and *COL6A1* rs35796750 were again shown to interact, with *COL12A1* rs970547, to modulate endurance running performance.

10.1.3 Tertiary Aim: The investigation of COL3A1 rs1800255, COL5A1 rs12722, COL6A1 rs35796750 and COL12A1 rs970547 in novel exercise-related phenotypes, namely exercise associated muscle cramps and rugby union playing level and position.

Due to the previously mentioned associations between collagen genes and musculoskeletal soft tissue injuries and other exercise-related phenotypes, this thesis also investigated *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 in novel exercise-related phenotypes, namely exercise associated muscle cramps (EAMC) and rugby union playing level and position.

A novel finding of this thesis was the identification of *COL5A1* rs12722 as a marker for a past history of EAMC. Specifically, the *COL5A1* rs12722 CC genotype was significantly over-represented in participants that reported never suffering from EAMC, when compared to participants that reported suffering from EAMC within 12 months prior to the events. A significant novel interaction was also identified between the *COL5A1* and *COL6A1* genes and risk of past history of EAMC. These results highlight the *COL5A1* rs12722 and *COL6A1* rs35796750 gene variants as risk factors for EAMC. The *COL3A1* rs1800255 and *COL12A1* rs970547 gene variants were not implicated in the aetiology of EAMC.

The novel findings of this thesis were that the *COL3A1* rs1800255 variant was independently associated with rugby union playing position. Professional rugby union forwards had a significantly lower rs1800255 GG genotype frequency distribution when compared to control participants. No independent associations were identified between *COL5A1* rs12722, *COL6A1* rs35796750 or *COL12A1* rs970547 and rugby union playing position. Despite this lack of independent associations, the *COL5A1* rs12722, *COL6A1* rs35796750 and *COL12A1* rs970547 gene variants were shown to interact to modulate individual inclination to a rugby union backs position.

10.1.4 Final Aim: To investigate the expression of the COL6A1 gene in individuals with known rs35796750 genotypes.

The *COL6A1* rs35796750 gene variant occurs near the branch site of intron 32 where the cytosine to thymine transition is proposed to cause aberrant splicing of the COL6A1 mRNA [191]. Similarly to the previously published functional studies on

COL5A1 rs12722 [107], it is important to determine the functional and biological mechanisms of how *COL6A1* rs3579750 may modulate exercise-related and other multifactorial phenotypes. This thesis therefore investigated the functional effects of *COL6A1* rs35796750.

This thesis showed for the first time that participants with the *COL6A1* rs35796750 TT and TC genotypes had significantly higher *COL6A1* gene expression when compared to participants with the CC genotype, suggesting that individuals with at least one *COL6A1* T allele have a higher type VI collagen gene expression. However analysis of *COL1A1* gene expression between the *COL6A1* genotype groups showed a similar *COL1A1*:*COL6A1* expression ratio in individuals with the *COL6A1* rs35796750 TC and CC genotypes, with a higher *COL1A1*:*COL6A1* ratio in the participants with a *COL6A1* TT genotype, suggesting that relative to type I collagen gene expression, type VI collagen might be higher in individuals with a *COL6A1* TT genotype. These novel results show a functional effect on *COL6A1* expression as a result of the rs35796750 gene variant.

As mentioned in each experimental chapter, there are limitations to the studies conducted in this thesis. Since these limitations may provide for false positive associations additional studies are required to replicate the results presented above. Despite this, the perspectives and future direction based on the results of this thesis will be discussed.

10.2 PERSPECTIVES

A summary of the novel findings of this thesis, as well as the results of previously published studies, for each gene with regard to the investigated musculoskeletal soft tissue injuries and other exercise-related phenotypes is presented in table 10.1. The contributions of these findings, together with the results of previously published studies, to understanding the role that each gene variant may play in the aetiologies of these phenotypes will be discussed.

Table 10.1. A summary of the specific genetic associations identified in this thesis. Both independent associations (genotype) and gene-gene interactions (haplotypes) are shown.

Investigated Phenotype	Associations	Investigated Gene Variants			
		COL3A1 **rs1800255 (G/A)**	***COL5A1*** **rs12722 (C/T)**	***COL6A1*** **rs35796750 (C/T)**	***COL12A1*** **rs970547 (A/G)**
ACL	**Genotype**	not associated	CC ♀ prev pub[1]	not associated	AA ♀ prev pub[2]
	Haplotype	nd	T ♀[3]	nd	A ♀[3]
Achilles tendinopathy	**Genotype**	not associated prev pub[4]	CC prev pub[5]	not associated	not associated prev pub[6]
	Haplotype	A	C	C	nd
ROM	**Genotype**	not associated	CC prev pub[7]	not associated	not associated
	Haplotype	not associated	C	T	not associated
Endurance Swim Performance[9]	**Genotype**	not associated	not associated	not associated	not associated
	Haplotype	G	C	T	not associated
Endurance Cycle Performance[9]	**Genotype**	not associated	not associated	TT	not associated
	Haplotype	not associated	not associated	not associated	not associated
Endurance Running Performance[9]	**Genotype**	not associated	TT prev pub[8]	not associated	not associated
	Haplotype	G	T	C	A
EAMC[9]	**Genotype**	not associated	CC	not associated	not associated
	Haplotype	not associated	C	T	not associated
Rugby Union[9]	**Genotype**	GG[10]	not associated	not associated	not associated
	Haplotype	not associated	T[11]	T[11]	G[11]

Green shading indicates that the allele or genotype is associated with a reduced risk of injury, except for ROM, endurance performance and rugby union where it represents an increased ROM or endurance performance and an over-representation in a position sub-group. Red shading indicates that the allele or genotype is associated with an increased risk of injury, except rugby union where it represents an under-representation in a position sub-group. Grey shading indicates no independent association. Prev pub, previously published. ♀, associations identified only in females. ACL, anterior cruciate ligament; ROM, range of motion; EAMC, exercise-associated muscle cramps; nd, not determined.

[1] independent association previously reported by Posthumus et al. [153].
[2] independent association previously reported by Posthumus et al. [154], but not independently associated in this thesis.
[3] T+A pseudo-haplotype also associated with increased risk of female ACL ruptures when a Polish cohort was analysed [137].
[4] no independent association previously reported by Saunders [170]
[5] independent association previously reported by Mokone et al. [131] and repeated by September et al. [179].
[6] no independent association previously reported by September et al. [181]
[7] independent association previously reported by Brown et al. [28;29].
[8] independent association previously reported by Posthumus et al. [151] and repeated by Abrahams et al. [1].
[9] only male participants were included in this study.
[10] Under-represented in the Rugby Union forwards sub-group in this thesis.
[11] Over-represented in the Rugby Union backs and outside backs sub-groups in this thesis.

10.2.1 The COL5A1 rs12722 Gene Variant

The associations between the rs12722 C/T variant within a functional region of the *COL5A1* 3'-UTR and these phenotypes, namely chronic Achilles tendinopathy, ACL ruptures in females, lower limb range of motion measurements and running endurance performance, have been previously described [1;2;28;29;40;131;151;153;156;179] and mechanisms for these associations have been proposed [41]. The novel findings that the *COL5A1* rs12722 CC genotype and C allele, as part of an inferred pseudo-haplotype, were associated with a reduced risk of a history of exercise-associated muscle cramps (EAMC) and that the T allele, as part of inferred pseudo-haplotypes, was associated with increased risk of a history of EAMC and with the backs (playing positions 9-15) rugby union positions are agreement with these previous findings and the mechanisms proposed for this variant [41] (Table 10.1). Additional evidence for the type V collagen genotype and exercise-related phenotype relationships hypothesis [41] was recently published, where the combined WW+CC genotypes of the *COL5A1* 3'-UTR rs14774622/rs55748801 and rs12722 was reported to be associated with reduced risk of the occupational over-use injury, carpal tunnel syndrome [32]. Although there is a growing body of evidence that hypothesizes that variants within the functional *COL5A1* 3'-UTR alter fibril architecture and structure and, thereby, directly or indirectly the mechanical properties of musculoskeletal soft tissue, there are studies that do not support the hypothesis (Chapter 2 Section 2.5.4) [58].

10.2.2 The COL3A1 rs1800255 Gene Variant

The *COL3A1* rs1800255 (G/A) variant was only independently associated with rugby union player position, while the A and G alleles were implicated in various collagen gene-gene interactions (Table 10.1). Specifically the rs1800255 GG genotype was shown to be under-represented in the forwards rugby union position sub-group, while the G allele was implicated in inferred pseudo-haplotype associations with increased endurance swimming and running performance (Table 10.1). In addition, the rs1800255 A allele was implicated in an inferred pseudo-haplotype associated with reduced risk of chronic Achilles tendinopathy. The non-synonymous *COL3A1* rs1800255 gene variant is proposed to be functional [94]. The guanine to adenine substitution results in an alanine to threonine change at position 698 of the COL3A1 peptide which is proposed to affect the tensile strength of type III collagen fibres [94]. The hydroxyl side chain of threonine is more hydrophilic than that of alanine and may result in distortion and unwinding of the collagen triple helix of type III collagen [94]. Interestingly, increased levels of type III collagen have also been shown to reduce fibril diameter in a dose dependent manner [110]. Although the exact implications that these changes to type III collagen may have on fibrillogenesis are not yet known, it is tempting to speculate that the AA genotype or A allele of rs1800255, containing the less favourable threonine residue at position 698, results in conformational changes in the triple helix of the type III collagen molecules, which when incorporated into the collagen fibril will alter its mechanical properties (Figure 10.1). Therefore, the presence of the more favourable GG genotype or G allele may grant endurance athletic benefits due to a more beneficial collagen fibril architecture and biomechanics, allowing for increased endurance swimming and running

performance (Figure 10.1). Although it is unclear how the incorporation of the type III collagen molecule with a “distorted and not properly wound” triple helix into the collagen fibril could be protective, the results of this thesis suggest that the G allele may increase the risk of musculoskeletal soft tissue injuries. This is highlighted by the association of the A allele with reduced risk of chronic Achilles tendinopathy (Figure 10.1). Furthermore, it cannot be discounted that this association is merely a consequence of the possible increased training volume, frequency and/or speed of endurance athletes who are able to perform better.

This proposed mechanism does not explain the under-representation of the “normal” GG genotype in rugby union forwards players, however previous studies have shown contrasting genetic association results between power and endurance sports [26]. For example, the GG genotype of the *PPARα* rs4253778 G/C variant was under-represented in power athletes (weightlifting and sprinting) and over-represented in endurance athletes (triathlons and cross-country skiing) when compared to apparently healthy unrelated controls [3].

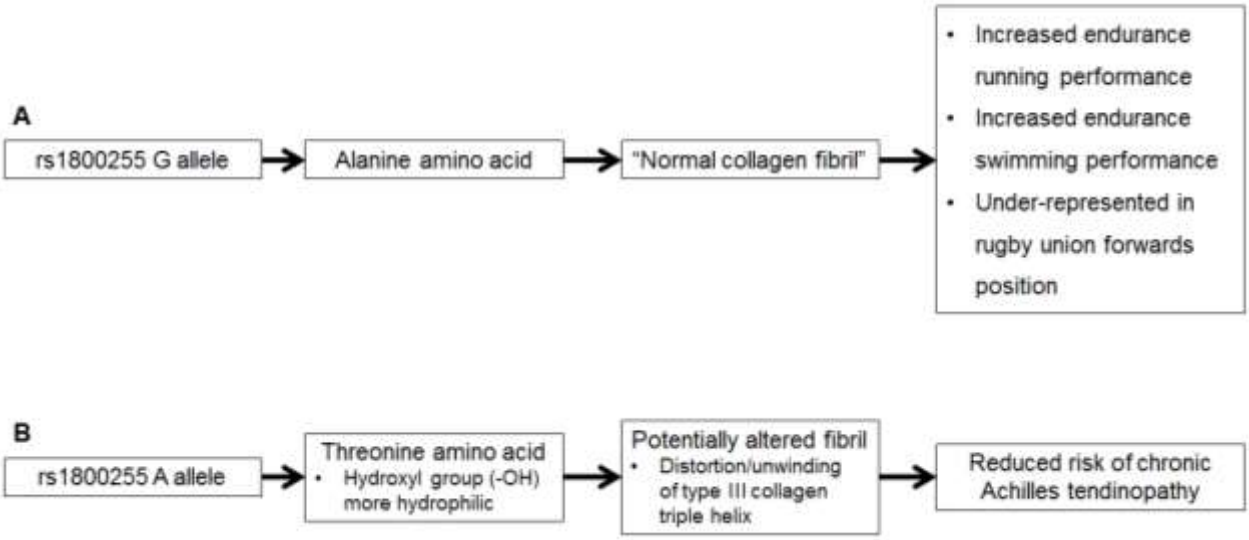


Figure 10.1. A proposed mechanism by which *COL3A1* rs1800255 G/A may alter the collagen fibril to modulate the aetiology of musculoskeletal soft tissue injuries and other exercise-associated phenotypes. (A) The rs1800255 G allele results in an alanine amino acid at position 698 of the COL3A1 peptide. This results in a "normal collagen fibril" with beneficial architecture and biomechanics, allowing for athletic benefits. (B) The rs1800255 A allele results in a more hydrophilic threonine amino acid at the same position. The resulting conformational changes, distortion and unwinding, of the triple helix of the type III collagen molecules, which when incorporated into the collagen fibril may alter its mechanical properties.

These varying genetic association results could also indicate that these associations may be due to another variant within the *COL3A1* gene or a neighbouring gene, such as *COL5A2*. Since *COL5A2* encodes for the α2(V) chain of type V collagen, it is possible that functional changes to the expression of *COL5A2* may affect type V collagen and the overall fibril in a similar manner as proposed for *COL5A1* [41]. Although less common, rare mutations within *COL5A2* have also been shown to cause the classical forms of EDS [122;124]. No individual variants within the

COL3A1 and downstream *COL5A2* genes have however to date been reported to associate with chronic Achilles tendinopathy [170]. In addition, no inferred haplotypes constructed from these *COL3A1* and/or *COL5A2* variants have been reported to associate with Achilles tendinopathy [170]. It is possible that associated variants within the *COL3A1*/*COL5A2* locus still need to be identified, and entire gene sequencing technologies and approaches could be considered for future studies.

Further studies are also required to confirm the proposed functional effects of rs1800255, as well as any other associated variant that might eventually be identified within the *COL3A1*/*COL5A2* locus, and how these variants may affect the biomechanical properties and architecture of type III collagen containing tissues. Furthermore, the associations identified with rs1800255 in this thesis were in male participants only and female participants should also be included in future studies. The possibility that rs1800255, and other *COL3A1* variants, is an important modulator of specific phenotypes also needs to be considered and investigated.

10.2.3 The COL12A1 rs970547 Gene Variant

This thesis did not identify any independent associations between the *COL12A1* rs970547 gene variant and investigated exercise-related phenotypes. However, novel inferred pseudo-haplotype associations were identified between the *COL12A1* rs970547 A/G gene variant and endurance running performance and rugby union position, respectively (Table 10.1). Specifically, the rs970547 A allele was associated with increased running performance, while the G allele was over-represented in the backs and outside backs rugby union position sub-groups (Table 10.1). Furthermore,

the *COL12A1* rs970547 AA genotype and A allele were previously associated with an increased risk of ACL rupture in females only [137;154]. It is interesting to note that, as reported for *COL3A1*, the opposite *COL12A1* alleles were associated with the endurance and power sports. Similarly, the same allele was associated with both endurance performance and risk of injury. ACL ruptures are not a common injury for endurance running athletes [20;61], however rugby union player are at higher risk since the sport involves high risk activities for ACL ruptures, including the cutting, pivoting and landing [20;61;167].

Although the function of rs970547 is unknown, bioinformatic functional analysis has shown that the glycine (rs970547 G allele) to serine (rs970547 A allele) change is potentially damaging to the α1(XII) chain and may modulate the function of type XII collagen [105]. Increasing levels of the N-terminal NC-3 domain of type XII collagen reduces inter-fibrillar connections, which may make the matrix surrounding these fibrils more pliable in the absence of cellular stress or more rigid if these domains are able to facilitate cell-mediated alignment and concentration of banded fibrils [133], and these changes to the matrix may then result in fibrils with reduced or larger diameters, respectively [133]. Therefore, it is possible that the rs970547 A allele may produce an isoform of type XII collagen which could alter fibril diameter and the biomechanical properties and architecture of the tissue. Future studies are required to test this proposal. Based on the proposed function of the rs970547 and the pattern of reported associations of this variant with the various phenotypes, a simple mechanism could not be proposed. The data does suggest that this *COL12A1* variant might modulate the effect that the *COL5A1* 3'-UTR variants have on some of the phenotypes (Table 10.1), such as ACL rupture (Chapter 3) [154], endurance

running performance (Chapter 6) and rugby union position (Chapter 8). It also cannot be excluded that these associations are due to other nearby polymorphisms within this gene. As stated in chapter 2, one other *COL12A1* variant (rs240736) was investigated for risk of chronic AT [181] and ACL ruptures [154], however no independent significant associations were observed. Future studies should investigate additional variants within this gene for associations with musculoskeletal soft tissue injuries and other exercise-associated phenotypes. Furthermore, additional studies should determine the mechanism through which types XII and V collagen may interact to modulate the aetiologies of these phenotypes.

10.2.4 The COL6A1 rs35796750 Gene Variant

Novel independent and haplotype associations were identified between the *CO6A1* rs35796750 C/T gene variant and musculoskeletal soft tissue injuries and other exercise-associated phenotypes (Table 10.1). Specifically, the TT genotype was independently associated with increased endurance cycling performance, while the T allele was implicated in inferred pseudo-haplotypes over-represented in the backs and outside backs rugby union position sub-groups and associated with increased endurance swimming performance and increased ROM, as well as, reduced risk of a history of EAMC, respectively (Table 10.1). In addition, the C allele was implicated in pseudo-haplotypes associated with increased endurance running performance, reduced risk of chronic Achilles tendinopathy and increased risk of a history of EAMC (Table 10.1).

Furthermore, this thesis showed for the first time that the rs35796750 C/T variant is functional (Chapter 9). Specifically, preliminary investigation showed that the rs35796750 TT genotype increases *COL6A1* gene expression, when compared to the CC genotype. Therefore, this thesis also proposed a genetic continuum for *COL6A1*, where common variants are associated with altered gene expression, similar to that proposed for the *COL5A1* gene in chapter 2 section 2.2 (Chapter 9, Figure 9.8) [41]. Mutations within a single copy of the *COL6A1* gene result in Bethlem myopathy and Ullrich congenital muscular dystrophy [109]. At the opposite end of the continuum, functional common variants within *COL6A1*, such as rs35796750, contribute to more complex and less severe phenotypes, and do not result in debilitating disorders (Chapter 9, Figure 9.8). A murine model has shown that *Col6a1 -/-* null mice have collagen fibrils with smaller diameters and greater density when compared to wildtype mice, accompanied by reduced tendon stiffness and maximal load [78]. Reduced *COL6A1* gene expression, as a result of the rs35796750 CC genotype, may therefore result in a similarly reduced collagen fibril diameter and increased fibril density, which in turn may reduce the stiffness and maximal load of the type VI collagen containing tissue including tendons, ligaments and muscle (Chapter 9, Figure 9.8). In contrast, the TT genotype may increase *COL6A1* gene expression resulting in increased collagen fibril diameter and a corresponding reduction in fibril density. A summary of the associations identified for the functional *COL6A1* rs3579750 are presented in figure 10.2.

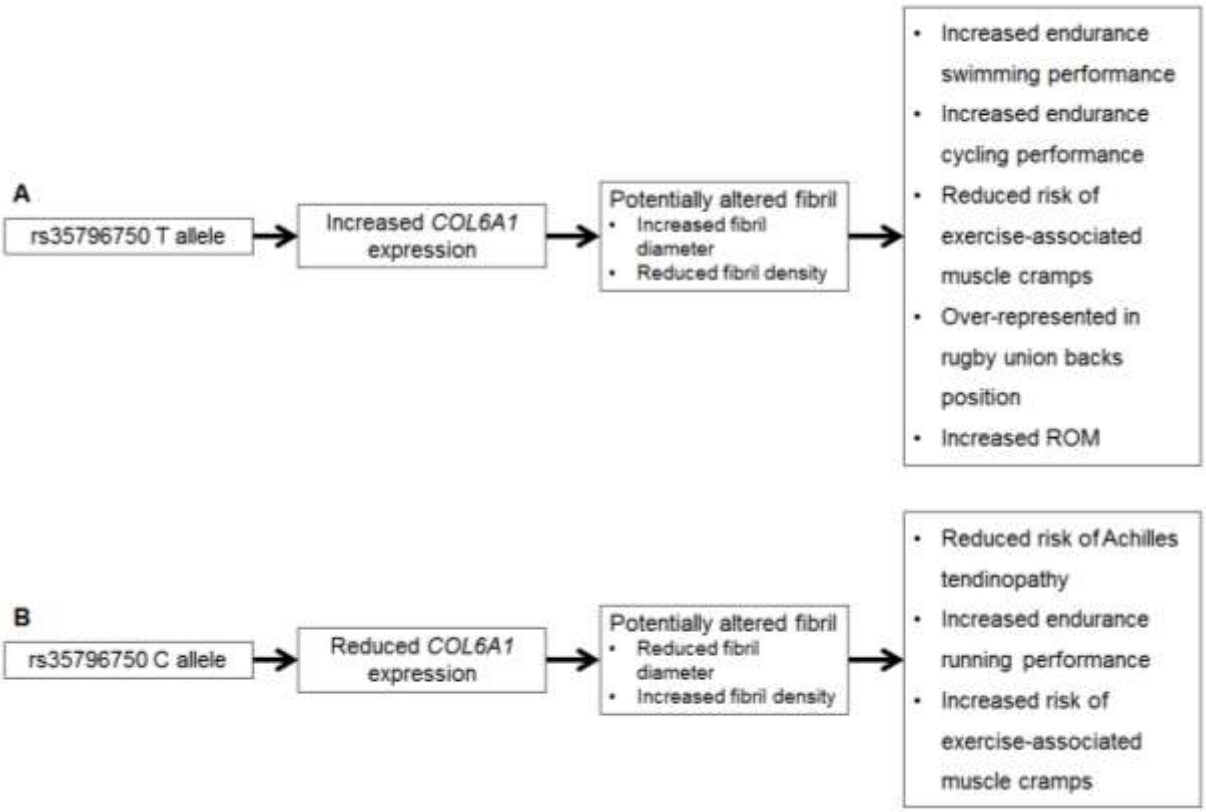


Figure 10.2. A proposed mechanism by which the functional *COL6A1* rs35796750 C/T may alter the collagen fibril to modulate the aetiology of musculoskeletal soft tissue injuries and other exercise-associated phenotypes. (A) The rs35796750 T allele results in increased *COL6A1* expression. This may result in an increased collagen fibril diameter and corresponding reduced density. (B) The rs35796750 C allele results in reduced *COL6A1* expression. This may result in a reduced collagen fibril diameter and corresponding increased density.

Interestingly, most of the associations identified for *COL6A1* rs35796750 may be explained by this mechanism (Chapter 9 Figure 9.8) and are consistent with the model outlined for type V collagen (Chapter 2 Figure 2.15) [41]. A summary of these similar and consistent findings, as well as contrasting and unique findings are presented in figure 10.3.

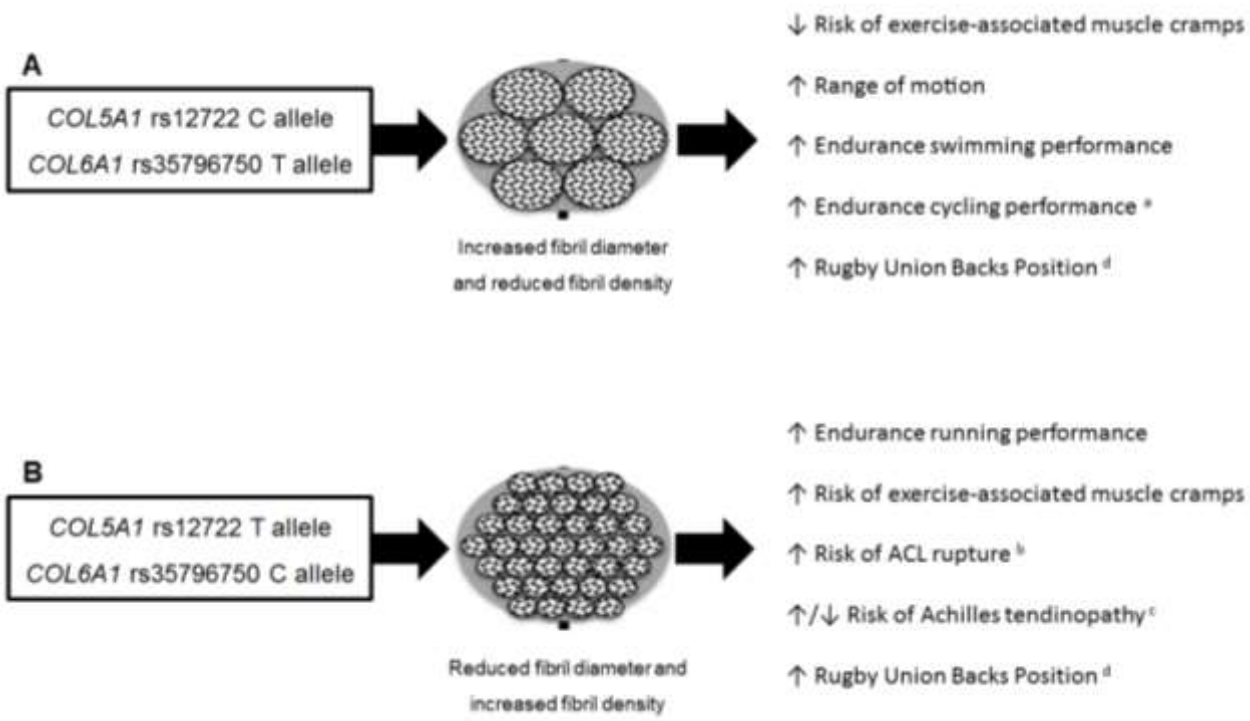


Figure 10.3. A summary of the association results for *COL5A1* rs12722 and *COL6A1* rs35796750, with regard to their proposed modulation of the collagen fibril. (A) The *COL5A1* rs12722 C and *COL6A1* rs35796750 T alleles are proposed to result in an increased fibril diameter and corresponding reduced density. (B) The *COL5A1* rs12722 T and *COL6A1* rs35796750 C alleles are proposed to result in a reduced fibril diameter and corresponding increased density. The up arrow (↑) indicates that the alleles are associated with an increased risk of injury, increased performance or increased range of motion. In the case of rugby union it represents an over-representation of the allele. The down arrow (↓) indicates that the alleles are associated with a reduced risk of injury.

[a] Only the *COL6A1* rs35796750 T allele was associated with endurance cycling performance.

[b] Only the *COL5A1* rs12722 T allele was associated with risk of ACL rupture.

[c] The *COL5A1* rs12722 T allele was associated with increased risk of chronic Achilles tendinopathy while the *COL6A1* rs35796750 C allele was associated with reduced risk.

[d] The *COL5A1* rs12722 T allele and contrasting *COL6A1* rs35796750 T allele were over-represented in the rugby union backs players.

Mutations within the *COL6A1* gene result in Bethlem myopathy [17;109] and Ullrich congenital muscular dystrophy [17;109;140] highlighting the importance of type VI collagen within skeletal muscle. Furthermore, mutations within *COL5A1* cause classic Ehlers-Danlos syndrome [124], suggesting a critical role for type V collagen in the extracellular matrix of musculoskeletal soft tissue. The *COL5A1* rs12722 (Chapter 2 Section 2.5.4) [107] and *COL6A1* rs35796750 (Chapter 9) variants have been shown to be functional. Specifically, these variants may alter the collagen fibril diameter and corresponding density (Figure 10.3). The changes to the fibril architecture as a result of *COL5A1* gene variants have been proposed to alter the soft tissue biomechanics, which may explain associations identified between these variants and musculoskeletal soft tissue injuries and other exercise-related phenotypes [41]. The aim of this thesis was to investigate collagen genes and variants, such as *COL6A1* rs35796750, that may have a similar effect on the collagen fibril architecture and therefore the tissue biomechanics. Most of the associations identified for *COL6A1* rs35796750 in this thesis support the model proposed by Collins and Posthumus (Figure 10.3) [41]. Interestingly, contrasting and unique associations were also identified. Specifically, contrasting association results were identified for risk of Achilles tendinopathy and for predisposition to the rugby union backs positions, while the *COL6A1* rs35796750 variant was not associated with ACL ruptures but was uniquely associated with endurance cycling performance (Figure 10.3). It is tempting to speculate that these contrasting and unique associations may be the result of the skeletal muscle specific nature of type VI collagen. These findings suggest a minor role for altered skeletal muscle biomechanics in ACL ruptures and predisposition to the rugby union backs positions, while skeletal muscle biomechanics may play a major role in the aetiology of

endurance cycling performance. Further research is required to test these hypotheses. Furthermore, it cannot be discounted that these contrasting and unique findings are the result of the multifactorial and polygenic aetiologies of these phenotypes.

10.2.5 The Multifactorial and Polygenic Nature of Musculoskeletal Soft Tissue Injuries and Other Exercise-Related Phenotypes

All the phenotypes investigated in this thesis, including ACL ruptures [67], chronic Achilles tendinopathy [128], joint range of motion [6;11;70;119;204], endurance performance [26;198], EAMC [174;176] and rugby union playing level and position [157;158;167;185], have complex multifactorial and polygenic aetiologies as outlined in figure 10.4.

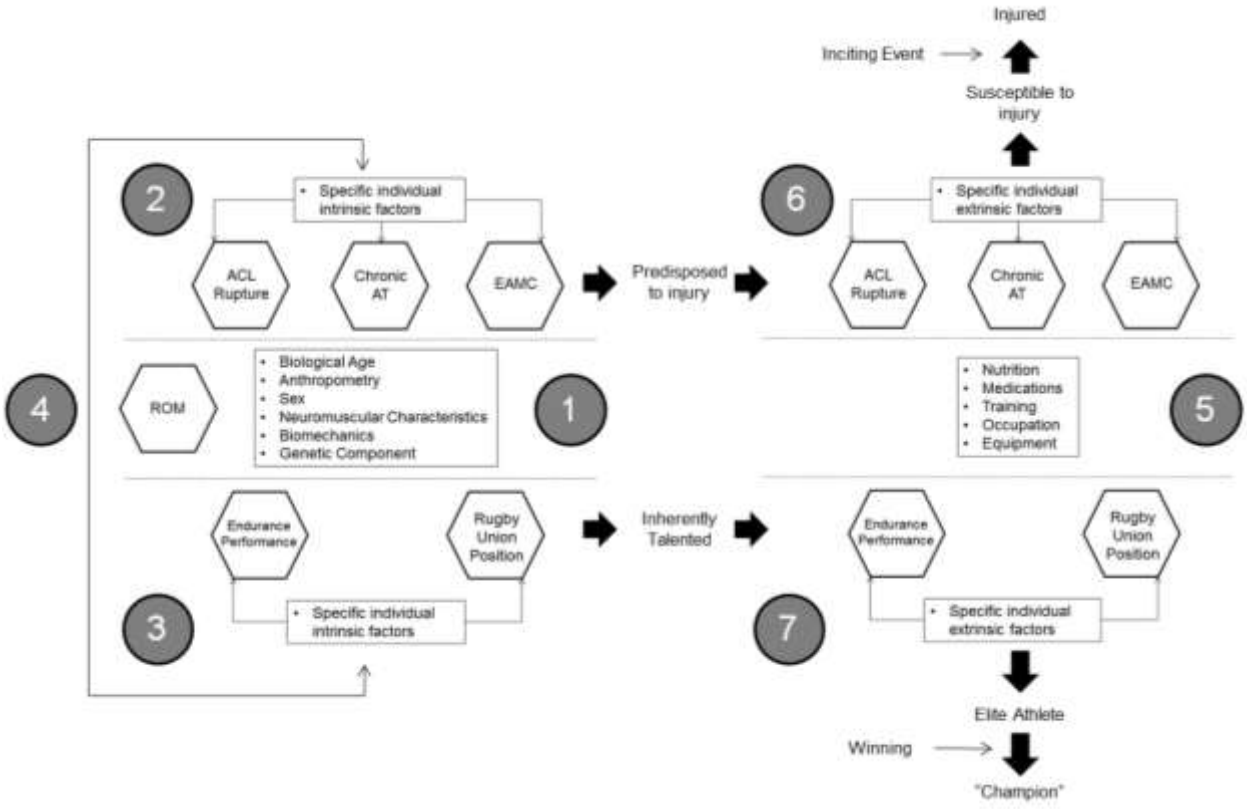


Figure 10.4. A summary of the multifactorial nature of the phenotypes investigated in this thesis, adapted from Tucker and Collins [198]. (1) The musculoskeletal soft tissue injuries and other exercise-related phenotypes investigated in this thesis have common predisposing intrinsic factors. These intrinsic factors include, among others, a common genetic component. Although range of motion (ROM) is a complex multifactorial phenotype it is also a common intrinsic factor in the aetiology of the other phenotypes investigated in this thesis. (2) In addition to the common intrinsic factors shown, the musculoskeletal soft tissue injuries have specific intrinsic factors, for example a previous injury of the same type, a previous injury to the same tissue and even specific genetic markers. (3) Similarly, other exercise-related phenotypes, including endurance performance and rugby union position, also have specific intrinsic factors such as specific genetic markers. (4) Musculoskeletal soft tissue injuries may be intrinsic factors in determining athletic ability, while athletic ability may be an intrinsic factor for musculoskeletal soft tissue injuries. (5) All of these phenotypes also have common extrinsic factors that may increase injury susceptibility or highlight elite athletes. (6) Each of the musculoskeletal soft tissue injuries also has specific extrinsic factors which increase an individual's susceptibility to that injury. An inciting event will then eventually result in injury. (7) Similarly, endurance performance and rugby union position also have specific extrinsic factors that highlight elite athletes. Winning an event or attaining international honours then results in "champions".

As shown in figure 10.4, the aetiologies of each of the phenotypes investigated in this thesis have numerous intrinsic factors. Furthermore, many of these intrinsic factors are common to all these phenotypes. In addition some of these multi-factorial phenotypes may even be intrinsic factors for others. For example ROM is a risk factor for musculoskeletal soft tissue injuries [208] and is also associated with running economy [77]. It is therefore difficult to determine which intrinsic factors may be directly causative to a phenotype and which are merely associated due to their indirect effects. These limitations are true for the genetic component as well. This is highlighted by the fact that none of the associated inferred pseudo-haplotypes had the same allele combinations (Table 10.1). Two possible reasons for these discrepancies are that; (i) the associations identified in this thesis may indicate direct causative factors or may be the result of causative loci in the same, or a neighbouring, gene. Or that (ii) these associations may be the result of indirect associations due to the overlap between the investigated phenotypes. Future research is required to determine the functional effects of all associated gene variants in order to better understand the molecular mechanisms, direct or indirect, through which their proteins might interact to modulate the aetiology of musculoskeletal soft tissue injuries and other exercise-related phenotypes.

This thesis has highlighted the role that collagens may play in the aetiology of exercise-related musculoskeletal soft tissue phenotypes. Collagen gene variants, namely *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs35796750 and/or *COL12A1* rs970547, have been identified as markers for the phenotypes investigated in this thesis, although in some cases independent associations were not identified. Although large effect sizes have been reported for some of the

common collagen gene variants studied to date, for example rs12722 [131;156;179], the expected effect sizes of common variants have often been described as approximately 1.1-1.5 and this may explain a lack of independent associations [4]. The sample sizes for this thesis were calculated based on previous findings for large effect sizes [131;156;179]. Future studies should investigate independent associations and gene-gene interactions, with biological relevance, within larger cohorts to overcome this limitation of common variants with small effect sizes.

The collagen genes chosen in this thesis all encode for collagens, which interact to have an impact on the structure and/or function of the collagen fibril. Therefore it may be hypothesised that minor changes to these collagens, as a result of the investigated common variants, may cumulatively modulate the collagen fibril thereby resulting in pseudo-haplotype markers for these phenotypes. This is evidenced by the association of multiple inferred pseudo-haplotypes with musculoskeletal soft tissue injuries and other exercise-associated phenotypes in this thesis, where independent associations were not identified for those variants. The biological significance of these inferred pseudo-haplotypes are however currently difficult to interpret.

10.3 FUTURE DIRECTIONS

The collagens investigated in this thesis have been identified in a number of tissues, including tendons, ligaments and skeletal muscle (Chapter 2, Table 2.5). It may be proposed that, due to the ubiquitous nature of these collagens, the number of functional variants with larger effect sizes within the genes that encode these

collagens will be limited since they could potentially result in a number of anomalous changes to all the tissues containing these collagens. These anomalous changes may not be tolerated during development or may result in severe disorders. Further complicating this issue is the fact that environmental factors may interact with these collagen variants in a tissue specific manner [38]. Therefore, the expression of these ubiquitous collagens may be different within specific tendon, ligament and skeletal muscle structures.

Unlike the collagens investigated in this thesis there are a number of known “lesser characterised” collagens which are not ubiquitous in nature, but rather are localised to particular tissues or to subdivisions within certain tissues (Table 10.2). The functions of these collagens within tendon, ligament and/or skeletal muscle tissues is largely unknown, however it is tempting to speculate that this expression specificity may result in reduced redundancy for specific functions performed by these collagens or that these collagens serve as redundancy for the ubiquitous collagens also expressed within these tissues. This lack of redundancy would mean that tolerated functional common variants within the genes that encode these “lesser characterised” collagens, which do not result in disorder but still alter the architecture and biomechanics of their containing tissues, are more likely to be independently associated with investigated phenotypes. In addition, the tissue specific nature of these “lesser characterised” collagens limits the confounding effects that environmental factors might have on their gene expression. Therefore the genes that encode these “lesser characterised” collagens are excellent candidates for future genetic association studies with musculoskeletal soft tissue injuries and other exercise-associated phenotypes.

Table 10.2. A summary of "lesser characterised" or minor collagens that have been identified within tendons, ligaments and the connective tissue structures of skeletal muscle.

Collagen Types	Tendon			Ligament	Skeletal Muscle		
	Midsub	MTJ	OTJ		Endo	Peri	Epi
Fibrillar							
XXVII					?	?	?
Non-fibrillar							
XIII		√			√		
XV					√	√	√
XIX					?	?	?
XX	?	?	?				
XXI					?	?	?
XXII		√					
XXIII	?	?	?				

The localisation of collagen types is indicated by "ticks". The question marks (?) show where a collagen type has been identified in that tissue but details were not given for a particular portion or subdivision. The tendon is shown as three separate portions; the midsubstance (Midsub), the myotendinous junction (MTJ) and the osteotendinous junction (OTJ). The three separate connective tissue subdivisions of skeletal muscle are shown; the endomysium (Endo), the perimysium (Peri) and the epimysium (Epi).

Type XXVII collagen is a fibrillar collagen that is mainly expressed in cartilage and at sites of transition from cartilage to bone but is also expressed in the skeletal muscle [23;74;144] (Table 10.2). The exact function of type XXVII within skeletal muscle is currently not known, however its specific localisation to this skeletal muscle tissue makes its encoding gene (*COL27A1*) an excellent candidate for genetic association studies to determine novel markers for musculoskeletal soft tissue injuries and other

exercise-associated phenotypes with skeletal muscle involvement, such as exercise-associated muscle cramps, endurance cycling performance and rugby union level and position. Interestingly, Saunders et al. [171] recently investigated the *COL27A1* and *TNC* genes as risk factors for chronic Achilles tendinopathy. No independent associations were identified for the investigated *COL27A1* rs946053, rs4143245, rs1249744 and rs753085 gene variants, however the rs946053 variant was implicated in a haplotype, with the *TNC* variants, associated with risk of chronic Achilles tendinopathy [171]. This study highlights that the *COL27A1* rs946053 may be a good candidate for future genetic association studies investigating musculoskeletal soft tissue injuries and other exercise-associated phenotypes.

Although little is known about their functions, a number of other "lesser characterised" non-fibrillar collagens, types XIII, XV, XIX, XXI and XXII collagen, are also specifically expressed within skeletal muscle and the myotendinous junction (Table 10.2). Type XV collagen is mainly derived from muscle cells and fibroblasts [68], and mice lacking type XV collagen suffer from skeletal myopathy highlighting the importance of this collagen within skeletal muscle [49]. In addition, types XIII and XXII collagen are expressed in the basement membrane of myotendinous junctions [72;99;189], and type XIII was also identified in the skeletal muscle endomysium [72;189]. Furthermore, types XIX and XXI collagen have also been identified within skeletal muscle [54;132]. As with type XXVII collagen the specific expression of these collagens makes their encoding genes excellent candidates for genetic association studies to determine novel markers for musculoskeletal soft tissue injuries and other exercise-associated phenotypes with skeletal muscle involvement.

Interestingly, the non-fibrillar types XX and XXIII collagen have been identified in tendons but not in ligaments or skeletal muscle [97;100] (Table 10.2). Type XX collagen has been identified as a FACIT while type XXIII collagen is a transmembrane domain collagen. The exact functions of these collagens have yet to be determined but high levels of type XXIII collagen in urine is a potential marker for lung cancer [188]. Further studies are required to determine the exact functions of these collagens and functional common variants within the *COL20A1* (encodes type XX collagen) and *COL23A1* (encodes type XXIII collagen) genes should be investigated for musculoskeletal soft tissue injuries and other exercise-associated phenotypes with tendon involvement, such as Achilles tendinopathy, endurance running performance, range of motion and rugby union playing level and position.

To date none of the “lesser characterised” collagens have been identified in ligaments (Table 10.2). This highlights the necessity for future studies to determine the exact functions of these “lesser characterised” collagens. Ligaments and tendons have a similar hierarchical structure, however their organisational and molecular structure are quite different [75]. As a result it is possible that tendon specific collagens, such as types XX and XXIII collagen, may also be expressed in ligaments, however it is also possible that ligament specific collagen types exist that remain to be identified. Therefore, at present, the “well characterised” collagens are the only source of novel collagen candidate genes for musculoskeletal soft tissue injuries and other exercise-associated phenotypes with ligament involvement, such as anterior cruciate ligament rupture, endurance running performance, range of motion and rugby union playing level and position. Both types IV and XIV collagens are expressed in ligaments (Chapter 2, Table 2.5), and therefore functional common

variants within their encoding genes, the *COL4A1-COL4A6* and *COL14A1* genes respectively, should be investigated for associations with these musculoskeletal soft tissue injuries and other exercise-associated phenotypes.

Finally, in addition to identifying novel variants within collagen genes which are associated with musculoskeletal soft tissue injuries and other exercise-associated phenotypes, future studies should determine the specific functions of these variants and the mechanisms through which they may have an effect on the collagen fibril and subsequent containing tissues. Of the collagen genes associated with musculoskeletal soft tissue injuries and other exercise-associated phenotypes only the *COL5A1* rs12722 [107] and *COL6A1* rs35796750 (Chapter 9) collagen gene variants have been investigated for their functional effects. Future studies should investigate the functional effects of those variants previously associated with musculoskeletal soft tissue injuries and other exercise-associated phenotypes, such as the *COL12A1* rs970547 and *COL27A1* rs946053 gene variants, and novel variants used in investigations for these phenotypes. Identifying the functional effects of these variants will provide understanding of the mechanisms through which these variants may affect the collagen fibril and thereby the collagen containing tissues. Furthermore, studies are also required to determine how the collagen molecules interact within the fibril and how changes to the fibril, resulting from collagen gene variants, may affect these interactions. These future studies would provide the data to develop accurate screening tools to determine individual susceptibility to musculoskeletal soft tissue injuries and other exercise-associated phenotypes.

10.4 CONCLUSION

In conclusion, this thesis identifies a number of novel associations between collagen gene variants and various musculoskeletal soft tissue injuries and other exercise-related phenotypes. Interestingly, the functional *COL6A1* rs35796750 variant was associated with most of the phenotypes investigated in this thesis. Furthermore, as initially hypothesised, the mechanism for these associations is consistent with that previously proposed for *COL5A1* rs12722 [41]. Future studies are required to investigate and understand the contrasting and unique results identified for *COL6A1* rs35796750. The *COL3A1* rs1800255 variant was associated with a number of phenotypes, while no additional independent associations were identified for *COL12A1* rs970547. In addition, varying gene-gene interactions were identified between the four variants and six phenotypes investigated in this thesis. The simplest interpretation of these findings is that the association between *COL5A1* rs12722 and the additional exercise-related phenotypes investigated in this thesis, namely exercise-associated muscle cramps and rugby union playing position, supports the previously described hypothesis that common variants within the type V collagen genes, especially *COL5A1*, are associated with seemingly unrelated musculoskeletal soft tissue phenotypes [41]. Furthermore, a similar genetic continuum may exist for type VI collagen. The results of this thesis however do not support the proposal that common variants, at least those investigated in this thesis, within the genes that encode other collagen types, namely III and XII, within or associated with the fibril are also associated with seemingly unrelated musculoskeletal soft tissue injuries. At best these other collagen variants appear to interact with *COL5A1* rs12722, and possibly *COL6A1* rs35796750, in a phenotype-

specific manner to modulate inter-individual variation. Further work is required to investigate the nature of these interactions and how they may modulate the aetiologies of these phenotypes. In addition, candidate variants within the genes that encode for “lesser characterised” collagens should be investigated in an attempt to identify additional independent markers for the phenotypes investigated in this thesis. These results highlight the complexity in identifying genetic, and other, markers for the inter-individual predisposition to musculoskeletal soft tissue injuries and other exercise-related phenotypes.

APPENDICES

(A) Ethical approval and recruitment forms

1. Approval letters from the human research ethics committee
2. Recruitment information sheets
3. Informed consent forms
4. Participant questionnaires
5. Diagnostic criteria forms

(B) Allelic Discrimination Methodology

1. PCR conditions for *COL1A1* rs1800012
2. PCR conditions for *COL5A1* rs12722
3. PCR conditions for *COL12A1* rs970547
4. PCR conditions for *COL3A1* rs1800255
5. PCR conditions for *COL6A1* rs35796750

(C) Supplementary Results

1. *COL3A1* rs1800255 and Achilles tendinopathy
2. *COL5A1* rs12722 and Achilles tendinopathy
3. *COL12A1* rs970547 and Achilles tendinopathy
4. COL5A1 rs12722 and Achilles tendinopathy
5. COL12A1 rs12722 and Achilles tendinopathy

(D) Quantitative Real-Time RT-PCR Methodology

(A) ETHICAL APPROVAL AND RECRUITMENT FORMS

1. Approval letters from the human research ethics committee

UNIVERSITY OF CAPE TOWN

Faculty of Health Sciences
Human Research Ethics Committee
Room E52-24 Groote Schuur Hospital Old Main Building
Observatory 7925
Telephone [021] 406 6338 • **Facsimile** [021] 406 6411
e-mail: shuretta.thomas@uct.ac.za

17 January 2013

HREC REF: 649/2012

Prof M Collins
Human Biology
Sport Science Institute

Dear Prof Collins

PROJECT TITLE: DETERMINING THE FUNCTIONAL ROLE OF VARIANTS WITHIN THE EXTRACELLULAR MATRIX GENES ON MUSCULOSKELETAL SOFT TISSUE INJURIES, USING PRIMARY HUMAN FIBROBLAST CELL LINES

Thank you for responding to the issues raised by the Faculty of Health Sciences Human Research Ethics Committee in your letter dated on 15th January 2013.

It is a pleasure to inform you that the HREC has **formally approved** the above-mentioned study.

Approval is granted for one year till the 30th January 2014

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/research/humanethics/forms)

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please quote the HREC. REF in all your correspondence.

Yours sincerely

pp T.Burgess

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN ETHICS
Federal Wide Assurance Number: FWA00001637.

Institutional Review Board (IRB) number: IRB00001938
This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Convention on Harmonisation Good Clinical Practice (ICH GCP) and Declaration of Helsinki guidelines.
The Human Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.

s.thomas

UNIVERSITY OF CAPE TOWN

FACULTY OF HEALTH SCIENCES
Human Research Ethics Committee

Annual Progress Report

Date	18 May 2012
REC REF Number	HREC REF 159/2011
Protocol number (if applicable) & Protocol title	The identification of genetic susceptibility loci underlying Achilles tendon pathology: A Repeat Study.
Principal Investigator	A/Prof Malcolm Collins
Department / Office Internal Mail Address	Exercise Science and Sports Medicine, Department of Human Biology, Sports Science Institute of South Africa

List of documentation

RESEARCH ETHICS COMMITTEE
2012 -05- 21
HEALTH SCIENCES FACULTY
UNIVERSITY OF CAPE TOWN

HREC office use only (FWA00001637; IRB00001938)			
☑ Approved	This serves as notification of annual approval, including all documentation described above.		
☐ Not approved	See attached comments.		
Type of review	☑ Expedited	☐ Full committee	
Expiry date	30 APRIL 2013		
Signature Chairperson of the HREC		Date	22/5/2012

UNIVERSITY OF CAPE TOWN

FACULTY OF HEALTH SCIENCES
Human Research Ethics Committee

FHS016: Annual Progress Report

HREC office use only (FWA00001637; IRB00001938)			
This serves as notification of annual approval, including any documentation described below.			
☑ Approved	Annual progress report		
☐ Not approved	See attached comments		
Expiry date	15 JUNE 2013		
Signature Chairperson of the HREC		Date	24/5/2012

RESEARCH ETHICS COMMITTEE
2012 -05- 23
HEALTH SCIENCES FACULTY
UNIVERSITY OF CAPE TOWN

Principal Investigator to complete the following:

1. Protocol information

Date	23 May 2012
HREC REF Number	139/2009
Protocol title	Identification of the genetic risk factors underlying [illegible] in all the population of South Africa
Protocol number (if applicable)	
Principal Investigator	Dr AV September
Department / Office Internal Mail Address	Department of Human Biology, ESSM, No 1 Boundary Road, Newlands

1.1 Does this protocol receive US Federal funding?	☐ Yes	X☐ No
1.2 Has sponsorship of this study changed? If yes, please attach a revised summary of the budget.	☐ Yes	X☐ No

2. List of documentation

3. Protocol status (tick ✓)

x☐	Open to enrolment

UNIVERSITY OF CAPE TOWN

FACULTY OF HEALTH SCIENCES
Human Research Ethics Committee

FHS016: Annual Progress Report

HREC office use only (FWA00001637; IRB00001938)			
This serves as notification of annual approval, including any documentation described below.			
☑ Approved	Annual progress report		
☐ Not approved	See attached comments		
Expiry date	15 JUNE 2013		
Signature Chairperson of the HREC		Date	23/5/2012

Principal Investigator to complete the following:

1. Protocol information

Date	22 May 2012
HREC REF Number	2008092- 092/2008
Protocol title	Genotype effects on range of motion (RoM) in healthy individuals.
Protocol number (if applicable)	
Principal Investigator	A/Prof Malcolm Collins
Department / Office Internal Mail Address	MRC/UCT Research Unit for Exercise Science & Sports Medicine, 3rd Floor, Sports Science Institute, Boundary Road, Newlands, 7700

1.1 Does this protocol receive US Federal funding?	☐ Yes	☒ No
1.2 Has sponsorship of this study changed? If yes, please attach a revised summary of the budget	☐ Yes	☒ No

RESEARCH ETHICS COMMITTEE
2012 -05- 23
HEALTH SCIENCES FACULTY
UNIVERSITY OF CAPE TOWN

2. List of documentation

3. Protocol status (tick ✓)

☐	Open to enrolment

UNIVERSITY OF CAPE TOWN

FACULTY OF HEALTH SCIENCES
Human Research Ethics Committee

FHS016: Annual Progress Report

HREC office use only (FWA00001637; IRB00001938)			
This serves as notification of annual approval, including any conditions detailed below.			
☐ Approved	Annual progress report		
☐ Not approved	See attached comments		
Expiry date	15 JUNE 2013		
Signature Chairperson of the HREC	Shirley Storey	Date	25-5-12

pp

Principal Investigator to complete the following:

1. Protocol Information

Date	23 May 2012
HREC REF Number	164/2006
Protocol title	The identification of genetic risk factors underlying anterior cruciate ligament injuries in all the populations of South Africa
Protocol number (if applicable)	
Principal Investigator	Dr Alison September
Department / Office Internal Mail Address	Department of Human Biology, ESSM, No 1 Boundary Road, Newlands

1.1 Does this protocol receive US Federal funding?	☐ Yes	X☐ No
1.2 Has sponsorship of this study changed? If yes, please attach a revised summary of the budget	☐ Yes	X☐ No

2. List of documentation

FHS006_164/2006 (genetic basis of ACL tears); project synopsis detailing rational for the new candidate genes for the 4 new students on the project and their names; a copy of the ethics approval of original project proposal (164/2006)

3. Protocol status (tick ✓)

x☐	Open to enrolment

RESEARCH ETHICS COMMITTEE
2012 -05- 25
HEALTH SCIENCES FACULTY
UNIVERSITY OF CAPE TOWN

UNIVERSITY OF CAPE TOWN

Health Sciences Faculty
Research Ethics Committee
Room E52-24 Groote Schuur Hospital Old Main Building
Observatory 7925
Telephone [021] 406 6338 • **Facsimile** [021] 406 6411
e-mail: lamees.emjedi@uct.ac.za

17 June 2010

HREC REF: 284/2010

A/Prof M Collins
Human Biology

Dear A/Prof Collins

PROJECT TITLE: ATHLETIC ABILITY DNA REPOSITORY AND DATABASE REGISTRY

Thank you for submitting your study to the Faculty of Health Sciences Research Ethics Committee for review.

It is a pleasure to inform you that the Ethics Committee has **formally approved** the above-mentioned study.

Approval is granted until 28 June 2011.

We note that data from the following studies are included in this registry:

1. Study:202/1999-Genetic markers for athletic ability, exercise performance and susceptibility to exercise induced injury
2. Study:005/2000- The Cape Town Ironman Triathlon 2000: fluid and sodium balance, predictors of performance, and medical consequences
3. Study:099/2001-South African Ironman 2001 Research Projects
4. Study:164/2001-Is the uncoupling protein 3 (ucp 3) gene associated with ultra-endurance performance?
5. Study:082/2004- Are the genes that encode for growth hormone or components of the kallikrein-kinin system associated with athletic ability
6. Study:185/2005-Normalisation of serum sodium concentrations in collapsed ultramarathon runners: oral versus intravenous fluid replacement
7. Study:429/2005-Genetic varients associated with athletic ability within South African male ultra-endurance athletes.

This registry is also related to a Master Registry under the auspices of Professor Malcolm Collins (HREC REF 284/2010)

Please submit an annual progress report if the research continues beyond the expiry date. Please submit a brief summary of findings if you complete the study within the approval period so that we can close our file.

lemjedi

Form FHS006: Protocol amendment

Instructions

- Forms to be downloaded from the Administrative Forms web page at http://web.uct.ac.za/depts/sapweb/forms/forms.htm
- All changes to the approved protocol must be reviewed and approved by the Human Research Ethics Committee (HREC) before implementation. See: 'Preparing an Amendment - Pointers for Researchers' on the Administrative Forms web page.

Submit Amendment and supporting documentation to:

Mrs Lamees Emjedi,
Research Ethics Committee
E 52 Room 24, Old Main Building, Groote Schuur Hospital, Observatory
Telephone: 27 21 406 6338
Fax: 27 21 406 6411
Email: Nosi.Tywabi@uct.ac.za

1. Protocol information

REC REF Number	(1) 284/2010, (2) 425/2005, (3) 002/2007, (4) 066/2009, (5) 092/2008, (6) 164/2006	
Title	(1) Athletic Ability DNA Repository and Database Registry (2) The research study at the Ironman 2006 triathlon, (3) The research study at the Ironman 2007 triathlon, (4) Two Oceans Ultra-Marathon 2009: Medical Consequences Following Endurance Sports, (5) The Genetic Effects on Flexibility Measurements, (6) Genetic Effects of Exercise-Induced Ligament Injury	
Principal Investigator	A/Prof Malcolm Collins and Mr Kevin O'Connell	
1.1 Does this protocol receive US Federal funding?		☐ No

2. Protocol status (tick ✓)

☐	Open to enrolment
☐	No participants have been enrolled
Y	Closed to enrolment (tick ✓)
Y	Research-related activities are ongoing
☐	Research-related activities are complete, long-term follow-up only
☐	Research-related activities are complete, data analysis only

3. Proposed changes will affect: (tick ✓ all the categories that apply)

	Protocol
☐	Study objectives, design (including investigator's brochure, clinical activities, study length)
☐	Study instruments, questionnaires, interview schedules
☐	Sample size

2. Recruitment information sheets

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: + 27 21 650 4561
Fax: + 27 21 686 7530

THE GENETIC BASIS OF EXERCISE-INDUCED CHRONIC TENDON PATHOLOGY

Although there is a high incidence of tendon overuse injuries as a result of participation in exercise and sporting activities, the cause(s) of these injuries are poorly understood. Some researchers have suggested that there is a genetic component to exercise-induced tendon injuries. In an attempt to determine whether there is a genetic basis for tendon pathology, we are interested in studying whether certain genes are associated with chronic tendinopathies. This project is being done in collaboration with the UCT/MRC Research Unit for Exercise Science and Sports Medicine within the Department of Human Biology and the Division of Human Genetics within the Department of Clinical Laboratory Sciences at the University of Cape Town.

You will be required to visit the Sports Science Institute of South Africa (SSISA) in Boundary Road, Newlands. During the visit, which should take at least 30 minutes, you will be asked to donate 5 ml (1 teaspoon) of a blood sample for DNA analysis. You will also be required to complete personal particulars, sporting details, medical history and stretching and warm up questionnaires. At a later stage, some participants will be asked to visit a doctor (radiologist) for a tendon scan at no cost to themselves.

All the information retrieved from this study will be treated with the strictest confidentiality and will be used only for scientific research purposes. Your name and personal particulars will not be released under any circumstances and all data will be analysed anonymously. Your DNA sample will be destroyed on completion of the study on the genetic basis of tendon pathology. You are also free to request that your DNA sample be destroyed before the completion of the study.

If you are part of the tendon pathology group, we would appreciate it if you could help us by recruiting two other people of same (or similar) age whom you know and who has trained without suffering any tendon injuries for the control group.

We will keep you informed about the outcomes of this study and look forward to working together with you. If you have any questions about this study, please feel free to contact us at:-

Dr. Malcolm Collins, PhD
(021) 650 4574
mcollins@sports.uct.ac.za

Prof. Martin Schwellnus, MBChB, MD
(021) 650 4576
mschwell@sports.uct.ac.za

Colleen Saunders, MSc student
(021) 650 4569
csanders@sports.uct.ac.za

Department of Human Biology
University of Cape Town
South Africa

16 July 2004

STUDY ON THE GENETIC BASIS OF TENDON PATHOLOGY

The Musculoskeletal Research Centre at La Trobe University in Melbourne, Australia, in collaboration with the UCT/MRC Research Unit for Exercise Science and Sports Medicine at the University of Cape Town (UCT) in Cape Town, South Africa, are currently studying the genetic basis of chronic tendon pathology in the Australian population.

Studies have suggested that some individuals have a genetic predisposition to tendon injury and that genes (those traits which you inherit form your parents), such as Type V collagen (COL5A1) and Tenascin C (TNC), which encode for important components of tendons are associated with tendon pathology.

The aim of this study is to determine whether the TNC, the COL5A1 or other similar genes are associated with chronic tendon pathology.

In order to participate in this study, you will be required to donate five millilitres of venous blood after giving written consent. The blood sample will be used for the extraction and analysis of genetic material (DNA). The extracted DNA will be sent to UCT in South Africa for analysis. The samples will be shipped to and analysed by UCT anonymously. The DNA will be genotyped (analysed) for variations (polymorphisms) within the *COL5A1*, *TNC* and other candidate genes believed to be associated with tendon pathology. You will be required to complete a number of questionnaires regarding personal particulars, sporting participation and medical history, as well as a stretching and warm up questionnaire. In addition, you will be required to visit a doctor (radiologist) at a later stage for a tendon(s) scan free of charge. All the information collected during the study will be treated with the strictest confidentiality and will only be used for scientific research purposes. Your name and personal particulars will not be released under any circumstances and all the data obtained will be analysed anonymously.

If you would like to participate in the study and/or obtain any additional information, please contact **Dr Jill Cook**, on phone: **9479 5789**, or e-mail: **J.Cook@latrobe.edu.au**.

Dear

RE: INVITATION TO PARTICIPATE IN A RESEARCH STUDY ON THE GENETIC PREDISPOSITION OF DEVELOPING ANTERIOR CRUCIATE LIGAMENT INJURIES

Dr Malcolm Collins and Prof Martin Schwellnus of the UCT/MRC Research Unit for Exercise Science and Sports Medicine within the Department of Human Biology at the University of Cape Town recently approached me about collaborating in a research study on the genetic basis of anterior cruciate ligament (ACL) injuries they are currently undertaking. It has been noted that certain individuals are predisposed to developing ACL injuries. The long term goal is to identify the various genes involved in ACL injuries. The researchers have asked me to inform you about this study and invite you to possibly participate in their research.

If you would like to participate in this study, you will be asked to visit the Sports Science Institute of South Africa in Newlands for a single visit to donate a five millilitre (a teaspoon) blood sample. After receiving written consent, the blood sample will be used for the extraction and analysis of genetic material (DNA). You will also be asked to complete personal particulars, sporting details, injury history, and medical history. The visit should not take longer than 30 minutes. All the information collected during the study will be treated with the strictest of confidentiality and will not be released under any circumstances. All the data obtained will be analysed anonymously.

If you would like to participate in the study and/or obtain any additional information please contact Michael Posthumus (details below). We would appreciate it if you let us know if you do not want to participate in the study so that we do not contact you again.

Michael Posthumus
Phone: (021) 650-4569
Cell: 073 190 5805
E-mail: mposthum@sports.uct.ac.za

Yours Sincerely

Dr Willem van der Merwe

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: + 27 21 650 4561
Fax: + 27 21 686 7530

THE GENETIC BASIS OF EXERCISE-INDUCED LIGAMENT INJURY

Although there is a high incidence of ligament injuries as a result of participation in exercise and sporting activities, the cause(s) of these injuries are poorly understood. Some researchers have suggested that there is a genetic component to exercise-induced ligament injuries. In an attempt to determine whether there is a genetic basis for ligament pathology, we are interested in studying whether certain genes are associated with ligament injuries. This project is being done in collaboration with the UCT/MRC Research Unit for Exercise Science and Sports Medicine within the Department of Human Biology of the University of Cape Town.

You will be required to visit the Sports Science Institute of South Africa (SSISA) in Boundary Road, Newlands. During the visit, which should take 15 minutes, you will be asked to donate 5 ml (1 teaspoon) of a blood sample for DNA analysis. You will also be required to complete personal particulars, sporting details, medical history and stretching and warm up questionnaires.

All the information retrieved from this study will be treated with the strictest confidentiality and will be used only for scientific research purposes. Your name and personal particulars will not be released under any circumstances and all data will be analysed anonymously. Your DNA sample will be destroyed on completion of the study on the genetic basis of ligament injury. You are also free to request that your DNA sample be destroyed before the completion of the study.

If you are part of the ligament pathology group, we would appreciate it if you could help us by recruiting two other people of same (or similar) age whom you know and who has trained without suffering any ligament or tendon injuries for the control group.

We will keep you informed about the outcomes of this study and look forward to working together with you. If you have any questions about this study, please feel free to contact us at:-

Dr Alison September, PhD
(021) 650 4559
alison.september@uct.ac.za

Dr Mike Posthumus, PhD
(021) 650 4572
michael.posthumus@uct.ac.za

Prof. Malcolm Collins, PhD
(021) 650 4574
malcolm.collins@uct.ac.za

Prof. Martin Schwellnus, MBChB, MD
(021) 650 4576
martin.schwellnus@uct.ac.za

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: +27 21 650 4561
Fax: +27 21 686 7530

THE GENETIC EFFECTS ON FLEXIBILTY MEASUREMENTS

Muscle and tendon injuries continue to make up a large proportion of the injuries reported by individuals participating in different sports. Many factors, such as incorrect training methods, muscle weakness, muscle strength imbalance, muscle fatigue and lower limb alignment abnormalities, are believed to be associated with these injuries. Reduced flexibility has also been cited as a risk factor for these injuries. Recent research within the UCT/MRC Research Unit for Exercise Science and Sports Medicine at the University of Cape Town (UCT) has suggested that an individual's flexibility is partly determined by their genetic make-up. We are therefore conducting this study to further determine whether specific genes are associated with flexibility.

To participate in this study you will be required to visit the Sports Science Institute of South Africa (SSISA) in Boundary Road, Newlands on two occasions. During the first visit, which should take at least 1 hour, you will be asked to donate 5 ml (1 teaspoon) of a blood sample for DNA analysis. You will also be required to complete personal particulars, sporting details, medical history and stretching and warm up questionnaires. Basic measurements such as height, weight and waist circumference will also be taken. Please do not perform any exercise training during the one day prior to your visit to SSISA. In addition please do not perform any stretching exercise two days prior to your visit.

Several flexibility measurements will also be done during the first visit. Electrodes from an EMG machine will be stuck to your skin (like a plaster) covering the thigh muscles in your legs. This will be used to record the electrical activity generated in your muscle during the flexibility test. A standard stretching procedure will then be used to increase the flexibility of your hamstring and shoulder. The flexibility measurements will be repeated immediately after the stretching procedure. Flexibility measurements will be done during the second visit which should last no longer than 15 minutes.

All the information retrieved from this study will be treated with the strictest confidentiality and will be used only for scientific research purposes. Your name and personal particulars will not be released under any circumstances and all data will be analysed anonymously. Your DNA sample will be destroyed on completion of the study on the genetic basis of tendon pathology. You are also free to request that your DNA sample be destroyed before the completion of the study.

You will not be reimbursed or compensated if you participated in this study. In addition you will not receive you personal genetic results. You will however be informed about the overall results of the study and your personal flexibility results.

The University of Cape Town (UCT) has an appropriate insurance policy to cover payment for any trial-related injury.

If you would like to participate in the study and/or obtain any additional information, please contact either:-

Caron-Jayne Miller
(021) 650 4569
MIlcar027@uct.ac.za

James Brown
(021) 650 4569
brwjam004@uct.ac.za

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: +27-21-650-4561 Fax: +27-21-686-7530

Dear

While our research group has been successful in identifying several genes associated with musculoskeletal soft tissue injuries and exercise-related traits, such as Achilles tendon ruptures or endurance running performance, much more work needs to be done in order to understand the relevance of these genetic associations at the functional level. Two of the associated genes, *COL5A1* and *COL6A1*, are involved in fibrillogenesis, the process by which collagen fibrils, the basic building block of tendons, ligaments and other connective tissues, are formed. It is proposed that naturally occurring inter-individual variations within these genes may alter normal fibrillogenesis and thereby affect the normal biomechanical properties of these tissues.

The aim of this study is to determine if these variants have functional effects on the genes in which they are identified, and the impact that they may have on the gene, since this would provide plausible mechanistic explanations for the associations found. In order to achieve this, tissue samples with known variants will be cultured and analysed. This study is being performed by Kevin O'Connell and Mary-Jessica Laguette who are both PhD students in the UCT/MRC Research Unit for Exercise Science and Sports Medicine, Department of Human Biology at the University of Cape Town.

You may be required to visit the Sports Science Institute of South Africa (SSISA) in Boundary Road, Newlands on two separate occasions. During the first visit, which should take no longer than 30 minutes, you will be asked to donate 5 ml (1 teaspoon) of a blood sample for DNA analysis. You will also be required to complete personal particulars, sporting participation, as well as personal and family medical history questionnaires. The DNA will be genotyped (analysed) for DNA sequence variations (polymorphisms) within candidate genes for musculoskeletal soft tissue injuries, such as the *COL5A1* and *COL6A1* genes.

You may be invited by the study investigators to participate in additional components of this study should your connective tissue genotypes match those required for the growth of primary fibroblast cell lines. You may be asked to visit the Sports Science Institute of South Africa (SSISA) in Boundary Road, Newlands on a second occasion in order to donate a single skin punch biopsy (3-4 mm), which will be used to culture primary skin fibroblast cell lines.

All the information retrieved from this study will be treated with the strictest confidentiality and will be used only for scientific research purposes. Your name and personal particulars will not be released under any circumstances and all data will be analysed anonymously. Your DNA and/or tissue sample will be destroyed on completion of the study on determining the functional role of variants within the extracellular matrix genes on

musculoskeletal soft tissue injuries, using primary human fibroblast cell lines. You are also free to request that your DNA and/or tissue sample be destroyed before the completion of the study.

You will not be reimbursed or compensated if you participated in this study. In addition you will not receive personal genetic results. You will however be informed about the overall results of the study and your personal flexibility results.

The University of Cape Town (UCT) has an appropriate insurance policy to cover payment for any trial-related injury.

We will keep you informed about the outcomes of this study and look forward to working together with you. If you have any questions about this study, please feel free to contact us at:-

Kevin O'Connell, PhD student
(021) 650 4569
kevin.oconnell@uct.ac.za

Mary-Jessica Laguette, PhD student
(021) 650 4569
nancylaguette@gmail.com

Prof Malcolm Collins, PhD
(021) 650 4574
malcolm.collins@uct.ac.za

3. *Informed consent forms*

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: + 27 21 650 4561
Fax: + 27 21 686 7530

GENETIC BASIS OF EXERCISE-INDUCED CHRONIC TENDON PATHOLOGY

INFORMED CONSENT

I, the undersigned, have been fully informed about the UCT/MRC Research Unit for Exercise Science and Sports Medicine within the Department of Human Biology and the Division of Human Genetics within the Department of Clinical Laboratory Sciences at the University of Cape Town's study on the genetic basis of exercise induced chronic tendon pathology. I have agreed to donate five millilitres of venous blood which will be used for the extraction and analysis of genetic material (DNA). I have also agreed to complete personal particulars, sporting participation, medical history, stretching and warm up questionnaires and understand that all the information that is collected during the study will be treated with the strictest confidentiality and will only be used for scientific research purposes. I also understand that my name and personal particulars will not be released under any circumstances and that all data will be analysed anonymously

If requested, I am also prepared to visit a doctor (radiologist) at a later stage for a tendon scan at no cost to myself (please delete this sentence if not applicable). If requested, I am also prepared to visit the SSISA for measurements to determine musculo-tendinous stiffness.

I agree to participate in the study and I have been informed that I will be free to withdraw from the study at any time if I so wish. I understand that my DNA sample will be destroyed on completion of the study on the genetic basis of tendon pathology. I also understand that I will be free to request that my DNA sample be destroyed before the completion of the study.

FULL NAME OF SUBJECT: _____

SUBJECT'S SIGNATURE: _____

DATE: _____

INVESTIGATOR: _____

INVESTIGATOR'S SIGNATURE: _____

13TH JULY 2004

LA TROBE UNIVERSITY

GENETIC BASIS OF TENDON PATHOLOGY

INFORMED CONSENT

I, (the participant), have been fully informed about this study on the genetic basis of tendon pathology to be conducted by the Musculoskeletal Research Centre at La Trobe University in Melbourne, Australia, in conjunction with the UCT/MRC Research Unit for Exercise Science and Sports Medicine at the University of Cape Town in Cape Town, South Africa.

I have agreed to donate five millilitres of venous blood, which will be used for the extraction and analysis of genetic material (DNA), and will be taken by a registered medical physician or nurse. The DNA will only be used for scientific research purposes relating to the genetic basis of tendon pathology only. I have also agreed to complete personal particulars, sporting participation, medical history, stretching and warm up questionnaires and understand that all the information that is collected during the study will be treated with the strictest confidentiality and will only be used for scientific research purposes. I also understand that all data will be analysed anonymously and my DNA sample will be destroyed on completion of the study.

I understand that the DNA extracted form the donated blood sample will be sent to UCT in South Africa for analysis. I understand that the DNA samples will be shipped to and analysed by UCT anonymously. I understand that the DNA will be genotyped (analysed) for variations (polymorphisms) within the COL5A1 and TNC genes, as well as other similar genes relating to the genetic basis of tendon pathology only.

In addition, I am also prepared to visit a doctor (radiologist) at a later stage for a tendon(s) scan at no cost to myself.

I understand that whilst there is no direct benefit to myself, if a genetic predisposition for tendinopathy can be established, then future generations will be able to establish their risk for this condition. This may allow better prevention and treatment options in the future.

I have read (or, where appropriate, have had read to me) and understood the information about this study, and any questions I have asked have been answered to my satisfaction. I agree to participate in the study, realising that I have the right to request that my DNA sample be destroyed at anytime

and, further, to demand that data arising from my participation is not used in the research project provided that this right is exercised within four weeks of the completion of my participation in the project. I agree that research data provided by me or with my permission during the project may be included in a thesis, presented at conferences and published in journals on the condition that neither my name nor any other identifying information is used.

Any questions regarding this project may be directed to the Chief Investigator: **Dr Jill Cook**, of the **School of Physiotherapy** on telephone number **9479 5789**.

If you have any complaints or queries that the investigator has not been able to answer to your satisfaction, you may contact the Ethics Liaison Officer, Human Ethics Committee, La Trobe University, Victoria, 3086, (ph: 03 9479 1443, e-mail: humanethics@latrobe.edu.au).

Name of Participant: _____

Signature: _____ **Date:** _____

Name of Researcher: _____

Signature: _____ **Date:** _____

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: +27 21 650 4561
Fax: +27 21 686 7530

GENETIC BASIS OF EXERCISE-INDUCED LIGAMENT INUJURY

INFORMED CONSENT

I, the undersigned, have been fully informed about the UCT/MRC Research Unit for Exercise Science and Sports Medicine within the Department of Human Biology of the University of Cape Town's study on the genetic basis of exercise induced chronic ligament pathology. I have agreed to donate five millilitres of venous blood or a Buccal mouthwash sample, which will be used for the extraction and analysis of genetic material (DNA). I have also agreed to complete personal particulars, sporting participation, medical history, stretching and warm up questionnaires and understand that all the information that is collected during the study will be treated with the strictest confidentiality and will only be used for scientific research purposes. I also understand that my name and personal particulars will be not released under any circumstances and that all data will be analysed anonymously.

I agree to participate in the study and I have been informed that I will be free to withdraw from the study at any time if I so wish. I understand that my DNA sample will be destroyed on completion of the study on the genetic basis of ligament pathology. I also understand that I will be free to request that my DNA sample be destroyed before the completion of the study.

I understand that the DNA will be genotyped (analysed) for variations (polymorphisms) within genes relating to the genetic basis of ligament injuries. I understand that whilst there is no direct benefit to myself, if a genetic predisposition for ligament injuries can be established, then future generations will be able to establish their risk for this condition. This may allow better prevention and treatment options in the future. I understand that I will receive the overall results of the study. I have read (or where appropriate, have had read to me) and understand the information about this study, and any questions I have asked have been answered to my satisfaction. I agree to participate in the study, realising that I have the right to request that my DNA sample be destroyed at any time. I agree that research data provided by me or with my permission during the project may be included in a thesis, presented at conferences and published in journals on the condition that either my name not any other identifying information is used.

FULL NAME OF SUBJECT: _____

SUBJECT'S SIGNATURE: _____

DATE: _____

INVESTIGATOR : _____

INVESTIGATOR'S SIGNATURE: _____

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: +27 21 650 4561
Fax: +27 21 686 7530

THE GENETIC EFFECTS ON FLEXIBILTY MEASUREMENTS

INFORMED CONSENT

I, _____ (the participant), have been fully informed about this study on the genetic basis of flexibility measurements to be conducted by the UCT/MRC Research Unit for Exercise Science and Sports Medicine at the University of Cape.

I have agreed to donate five millilitres of venous blood, which will be used for the extraction and analysis of genetic material (DNA), and will be taken by a phlebotomist. The DNA will only be used for scientific research purposes relating to the genetic basis of flexibility measurements. I agree to my height, weight and waist circumference measured. I have also agreed to complete questionnaires relating to personal particulars, sporting participation, medical history, stretching and warm up exercise and understand that all the information that is collected during the study will be treated with the strictest confidentiality and will only be used for scientific research purposes. I also understand that all data will be analysed anonymously and my DNA sample will be destroyed on completion of the study. I understand that the DNA will be genotyped (analysed) for variations (polymorphisms) within the *COL5A1* and other similar genes relating to the genetic basis of flexibility measurements.

I am also prepared to have the flexibility of my upper and lower limbs measured. I am prepared to have my upper leg prepared for and have EMG electrodes attached to measure the activity in my muscles during the lower limb flexibility test. I have agreed to allow the researchers to do a standard stretching procedure to increase my hamstring and shoulder flexibility.

The potential risks associated with blood collection technique from the ante-cubital veins are: infection, delayed healing, haematoma, physical pain, mental discomfort and injury to a nerve or a vessel. These risks are small and will be minimized by the use of trained phlebotomists, use of sterile techniques and the use of disposable, single use materials.

I understand that whilst there is no direct benefit to myself, a genetic predisposition for flexibility can be established. I have read (or, where appropriate, have had read to me) and understood the information about this study, and any questions I have asked have been answered to my satisfaction. I agree to participate in the study, realising that I have the right to request that my DNA

sample be destroyed at anytime and, further, to demand that data arising from my participation is not used in the research project provided that this right is exercised within four weeks of the completion of my participation in the project. I agree that research data provided by me or with my permission during the project may be included in a thesis, presented at conferences and published in journals on the condition that neither my name nor any other identifying information is used.

Any questions regarding this project may be directed to the Principle Investigator: **Prof Malcolm Collins** on telephone number **021 650 4574** or e-mail **malcolm.collins@uct.ac.za**.

If you have any complaints or queries that the investigator has not been able to answer to your satisfaction, you may contact the Faculty of Health Sciences Human Research Ethics Committee at the University of Cape Town **Prof Marc Blockman** on telephone number **021 406 6452**.

Name of Participant: _____

Signature: _____ **Date:** _____

Name of Researcher: _____

Signature: _____ **Date:** _____

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: +27 21 650 4561
Fax: +27 21 686 7530

DETERMINING THE FUNCTIONAL ROLE OF VARIANTS WITHIN THE EXTRACELLULAR MATRIX GENES ON MUSCULOSKELETAL SOFT TISSUE INJURIES, USING PRIMARY HUMAN FIBROBLAST CELL LINES.

SKIN PUNCH BIOPSY INFORMED CONSENT

I, the undersigned, have been fully informed about the UCT/MRC Research Unit for Exercise Science and Sports Medicine within the Department of Human Biology of the University of Cape Town's study to identify the functional role of variants within the extracellular matrix (connective tissue) genes on musculoskeletal soft tissue injuries, using primary human fibroblast cell lines. A single skin punch biopsy (3-4 mm), will be used to culture primary skin fibroblast cell lines. A fibroblast is a type of cell that synthesizes the structural framework (extracellular matrix and collagen) for tissues. The punch biopsy will be done by a trained medical practitioner. All the information that is collected during the study will be treated with the strictest confidentiality and will only be used for scientific research purposes. Names and personal particulars will be not released under any circumstances and all data will be analysed anonymously.

I agree to participate in the study and to donate a single skin punch biopsy as described above. I have been informed that I will be free to withdraw from the study at any time if I so wish. The tissue samples will be destroyed on completion of the study to identify the functional role of variants within the extracellular matrix (connective tissue) genes on musculoskeletal soft tissue injuries, using primary human fibroblast cell lines. I also understand that I will be free to request that my tissue sample be destroyed before the completion of the study.

The potential risks associated with the skin punch biopsy technique from the forearm are: infection, delayed healing, physical pain, mental discomfort and injury to a nerve or a vessel. These risks are small and will be minimized by the use of trained medical practitioner, use of sterile techniques and the use of disposable, single use materials.

The tissue sample will be cultured and analysed using a variety of molecular techniques to determine the functional effects of DNA sequence variations (polymorphisms) within candidate genes for musculoskeletal soft tissue injuries, such as the *COL5A1* and *COL6A1* genes. Any remaining tissue will be discarded appropriately.

There is no direct benefit to myself however, this may allow better prevention and treatment options for musculoskeletal soft tissue injuries in the future. I understand that I will receive the overall results of the study. I

have read (or where appropriate, have had read to me) and understand the information about this study, and any questions I have asked have been answered to my satisfaction. I agree to participate in the study, realising that I have the right to request that my tissue sample be destroyed at any time. I agree that research data provided by me or with my permission during the project may be included in a thesis, presented at conferences and published in journals on the condition that neither my name nor any other identifying information is used.

Any questions regarding this project may be directed to the Principle Investigator: **Prof Malcolm Collins** on telephone number **021 650 4574** or e-mail **malcolm.collins@uct.ac.za**.

If you have any complaints or queries that the investigator has not been able to answer to your satisfaction, you may contact the Faculty of Health Sciences Human Research Ethics Committee at the University of Cape Town **Prof Marc Blockman** on telephone number **021 406 6452**.

FULL NAME OF SUBJECT: _____

SUBJECT'S SIGNATURE: _____

DATE: _____

INVESTIGATOR : _____

INVESTIGATOR'S SIGNATURE: _____

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: +27 21 650 4561
Fax: +27 21 686 7530

DETERMINING THE FUNCTIONAL ROLE OF VARIANTS WITHIN THE EXTRACELLULAR MATRIX GENES ON MUSCULOSKELETAL SOFT TISSUE INJURIES, USING PRIMARY HUMAN FIBROBLAST CELL LINES.

BLOOD COLLECTION AND DISCARDED TISSUE INFORMED CONSENT

I, the undersigned, have been fully informed about the UCT/MRC Research Unit for Exercise Science and Sports Medicine within the Department of Human Biology of the University of Cape Town's study to identify the functional role of DNA (genetic material) sequence variants within the extracellular matrix (connective tissue) genes on musculoskeletal soft tissue injuries, using primary human fibroblast cell lines. A fibroblast is a type of cell that synthesizes the structural framework (extracellular matrix and collagen) for tissues. Participants are required to donate five millilitres of venous blood or a Buccal mouthwash/swab sample, which will be used for the extraction and analysis of DNA. The blood sample will be taken from a forearm (ante-cubital) vein by a nurse, physician or phlebotomist. The DNA will be genotyped (analysed) for DNA sequence variations (polymorphisms) within candidate genes for musculoskeletal soft tissue injuries, such as the *COL5A1* and *COL6A1* genes, to investigate their effects on connective tissue, such as tendons and ligaments. Participants are required to complete personal particulars, sporting participation, as well as personal and family medical history questionnaires. All the information that is collected during the study will be treated with the strictest confidentiality and will only be used for scientific research purposes. Names and personal particulars will be not released under any circumstances and all data will be analysed anonymously.

I agree to participate in the study. I agree to donate a blood sample for the extraction and analysis of my DNA. I agree to complete all study related questionnaires as outlined above. I have been informed that I will be free to withdraw from the study at any time if I so wish. I understand that my DNA sample and/or tissue sample will be destroyed on completion of the study to identify the functional role of variants within the extracellular matrix (connective tissue) genes on musculoskeletal soft tissue injuries, using primary human fibroblast cell lines. I also understand that I will be free to request that my DNA sample and/or tissue sample be destroyed before the completion of the study.

The potential risks associated with blood collection technique from the forearm (ante-cubital) veins are: infection, delayed healing, haematoma, physical pain, mental discomfort and injury to a nerve or a vessel. These risks are small and will be minimized by the use of trained phlebotomists, use of sterile techniques and the use of disposable, single use materials.

The DNA will be genotyped (analysed) for DNA sequence variations (polymorphisms) within candidate genes for musculoskeletal soft tissue injuries, such as the *COL5A1* and *COL6A1* genes.

If I am undergoing surgery, I agree to allow any of my tissue that is excised during surgery and to be discarded for incineration, to be collected for this study. This collected tissue will be used to culture primary fibroblast cell lines, which will be used in a number of molecular techniques to identify the functional role of variants within candidate genes for musculoskeletal soft tissue injuries, such as the *COL5A1* and *COL6A1* genes. Any remaining tissue will be discarded appropriately.

I give permission to be contacted by the study investigators to be invited to participate in additional components of this study should my connective tissue genotypes match those required for the growth of primary fibroblast cell lines.

I understand that there is no direct benefit to myself however, this may allow better prevention and treatment options for musculoskeletal soft tissue injuries in the future. I understand that I will receive the overall results of the study. I have read (or where appropriate, have had read to me) and understand the information about this study, and any questions I have asked have been answered to my satisfaction. I agree to participate in the study, realising that I have the right to request that my DNA sample and/or tissue sample be destroyed at any time. I agree that research data provided by me or with my permission during the project may be included in a thesis, presented at conferences and published in journals on the condition that neither my name nor any other identifying information is used.

Any questions regarding this project may be directed to the Principle Investigator: **Prof Malcolm Collins** on telephone number **021 650 4574** or e-mail **malcolm.collins@uct.ac.za**.

If you have any complaints or queries that the investigator has not been able to answer to your satisfaction, you may contact the Faculty of Health Sciences Human Research Ethics Committee at the University of Cape Town **Prof Marc Blockman** on telephone number **021 406 6452**.

FULL NAME OF SUBJECT: _____

SUBJECT'S SIGNATURE: _____

DATE: _____

INVESTIGATOR : _____

INVESTIGATOR'S SIGNATURE: _____

2000/2001 South African Ironman triathlon informed consent form.

INFORMED CONSENT

I, _____, have read and fully understand the details of the study and hereby give consent to participate in the trial to be conducted by the MRC/UCT Research Unit for Exercise Science and Sports Medicine of the Department of Human Biology, University of Cape Town, at the Cape Town IRONMAN Triathlon to be held in Gordons Bay on 31 March 2001.

I agree to participate in the study and understand the commitments required of me. I understand that on completion or early termination of the triathlon, I will enter the medical tent for a full medical examination and will allow the medical staff to administer medical treatment that is in the best interest of my health and/or recovery. I am aware that I may withdraw from the trial at any time.

Name of triathlete: _____

Signature of triathlete _____

Date _____

Name of investigator _____

Signature of Investigator _____

Date _____

5

06/2009

INFORMED CONSENT FORM

I, _____, agree voluntarily to participate in the following components **(DELETE THOSE COMPONENTS YOU DO NOT AGREE TO PARTICIPATE IN)** of the UCT/MRC Research Unit for Exercise Science and Sports Medicine's, University of Cape Town, research project titled:-

1. Causes of Exercise Associated Muscle Cramping (EAMC) in ultra-marathon runners
2. The genetic basis for common running related injuries, particularly the response of soft tissues (Achilles tendon) to repeated forces during running (mechanical loading)
3. Pre-race predictors (including training parameters, medical history, medication use, and psychological characteristics (traits) of medical complications that may occur in runners during and immediately after the Two Oceans ultra-marathon
4. Factors associated with pre-race and post-race (up to 10 days) respiratory tract symptoms in runners
5. Brain and nervous system tiredness (Neural Fatigue) following the Two Oceans ultra-marathon

I understand that my participation in this research project has no direct benefits to me during the Two Oceans 2009 competition. However, I understand that my participation in the research project will advance the medical and scientific knowledge related to endurance sports. Therefore, information gathered through my participation in this project could advance the future medical care, training advice and performance of endurance athletes.

I have read the subject information sheets and the following procedures and concepts have been explained to me in full:

(DELETE THOSE COMPONENTS YOU DO NOT AGREE TO PARTICIPATE IN)

1. **Completion of a questionnaire: (all components)**

The completion of personal details, racing, training, equipment use, medical, supplement use, fluid use and lifestyle history questionnaires are not associated with any risk. Completion of self-rated behavioural questionnaires has not previously been shown to be associated with risk. A potential risk is that people who have experienced significant past trauma will find questionnaires on this uncomfortable. The questions within the behavioural questionnaires are asking are about personality characteristics (temperament) and none of the scales are directed at picking up psychological abnormalities (psychopathology). Any personal identification of subjects (names and surnames), questionnaire data and other clinical data (paper and electronic) will be kept

confidential, will be kept secure, and will not be made available to any party other than the research team without the consent of the individual subjects.

I agree that the all the questionnaire information, my performance during the Two Oceans marathon, together with all the other data collected from the various components of this trial may be used to answer scientific questions about the medical conditions, physiological responses and measures of performance associated with the participation in and completion of an ultramarathon.

2. **Blood sample collection for re- serum creatine kinase (marker of muscle damage) levels (only for the cramps component)**

I have agreed to donate 5 milliliters (1 teaspoon) of venous blood during registration. The sample will be used to measure my levels of a muscle enzyme that is released if muscle is damaged (serum creatine kinase levels). The potential risks associated with blood collection technique from the veins on my arm (ante-cubital veins) are: infection, delayed healing, blood clot (haematoma), physical pain, mental discomfort and injury to a nerve or a vessel. These risks are small and will be minimized by the use of staff that is trained to take blood samples (trained phlebotomists), use of sterile techniques and the use of disposable, single use materials. The risk of decreased performance as a result of blood collection will be reduced by not subjecting any participant to the collection of a blood volume exceeding 15 ml prior to the race.

3. **Measurement of Flexibility : (only for the cramps component)**

I have agreed to undergo measurements of my lower limb flexibility. I understand that the two tests are the straight-leg raise test, and the sit-and-reach test. In both these test my limbs will be relaxed and electrodes to measure muscle activity will be attached to my skin (back of the thigh). A tester who is experienced in administering these test, will then perform the tests as follows: 1) my leg will be raised (with the knee straight) until it feels "tight" and a muscle activity signal is registered on a machine that measures electrical signals from the muscles (electromyographic machine - EMG) - the degree to which my leg is lifted will be measured. 2) I will be asked to sit on the floor and then reach forward with both hands until it feels tight – the distance from my fingertips to my toes will then be measured. These are standard tests that are used daily in clinics and are associated with minimal risk. The only risk is to overstretch, but I will have the freedom to stop the test at any time if the stretch becomes uncomfortable.

4. **Recording of heart rate variability during stroop test: (only for the management of pain components)**

The stroop test is a simple, computer based test (similar to a computer game). The mental concentration that is required for the test is relevant for the data collection and not the outcome of the test. There is no risk associated with the recording of the heart rate variability

5. **Soft tissue diagnostic ultrasound examination: (only for the Achilles tendon component)**

I understand that I will be subjected to a test where gel is applied to my skin and a special machine is used to see the Achilles tendon underneath the skin (soft tissue diagnostic ultrasound examination of my Achilles tendons) during the registration period, on completion of the race, and if possible 6 weeks after the race at a medical facility close to my home. I understand that I will not receive any direct financial compensation to attend this centre for the ultrasound, but that the investigation will be free of charge. I understand that these investigations are not associated with any risk, and will be performed by a trained radiologist.

6. **Blood sample collection for genetic studies: (only for the genetics component)**

At one of the pre-race facilities or at race registration, I have agreed to donate ten milliliters (2 teaspoons) of venous blood. The sample will be used for the extraction and analysis of genetic material (DNA).

The potential risks associated with blood collection technique from the veins on my arm (ante-cubital veins) are: infection, delayed healing, blood clot (haematoma), physical pain, mental discomfort and injury to a nerve or a vessel. These risks are small and will be minimized by the use of staff that is trained to take blood samples (trained phlebotomists), use of sterile techniques and the use of disposable, single use materials. The risk of decreased performance as a result of blood collection will be reduced by not subjecting any participant to the collection of a blood volume exceeding 15 ml prior to the race.

The genetic material that is extracted from my blood (DNA) will only be used for scientific research purposes relating to the genetic basis of (1) athletic ability, (2) physiological response to (3) medical complications during ultra-endurance events. I have also agreed to complete personal particulars, training, sporting, measures of my inherent behaviour and responses (behavioural endophenotypes) and medical questionnaires and understand that all the information that is collected during the study will be treated with the strictest confidentiality and will only be used for scientific research purposes.

Questionnaire and other clinical data (paper and electronic) will be kept confidential, will be kept secure, and will not be made available to any party other than the research team without the consent of the individual subjects. I also understand that all data will be analysed without revealing any of my personal details (anonymously) and my DNA sample will be destroyed on completion of the study.

I understand that some of the DNA extracted from the donated blood sample will be sent to the Cyprus Institute of Neurology and Genetics in Cyprus for analysis. I understand that the DNA samples will be shipped to and analysed in Cyprus anonymously. I understand that the DNA will be genotyped (analysed) for variations (polymorphisms) within genes relating to the genetic basis of athletic ability, tendon and ligament overuse injuries and dysnatraemia during ultra-endurance events only.

I understand that whilst there is no direct benefit to myself, if a genetic predisposition for (1) athletic ability, (2) physiological response to and (3) medical complications during ultra-endurance events can be established, then future generations will be able to establish their risk for this condition. This may allow better prevention and treatment options in the future. I understand that I will receive the overall results of the study.

I have read (or, where appropriate, have had read to me) and understood the information about this study, and any questions I have asked have been answered to my satisfaction. I agree to participate in the study, realising that I have the right to request that my DNA sample be destroyed at anytime. I agree that research data provided by me or with my permission during the project may be included in a thesis, presented at conferences and published in journals on the condition that neither my name nor any other identifying information is used.

I have read the preceding subject information sheet and understand the testing procedures outlined therein. I understand any accompanying risks and discomforts. Knowing these risks and discomforts and having had the opportunity to pose questions answered to my satisfaction, I hereby consent to participate in this study. I understand that I may withdraw from this study at any time without further question. I have been informed that the individual data derived from my participation in these protocols will remain confidential. I understand that the medical staff and the research team have professional medical insurance.

Name of the tri-athlete: _____

Signature of tri-athlete _____

Date: _____

Name of investigator: _____Prof Martin Schwellnus____

Signature of Investigator: _____

Date: _____

SPORTS SCIENCE INSTITUTE OF SOUTH AFRICA

Appendix B

University Of Cape Town

SUBJECT INFORMED CONSENT FORMS:

COMPONENTS OF THE RESEARCH STUDY TO BE CONDUCTED AT THE 2006 IRONMAN TRI-ATHLON IN PART ELIZABETH

The research study at the 2006 Ironman triathlon, comprise of six components. Detailed informed consent forms for each component of the study are as follows:

INFORMED CONSENT FORM

Component 1: A study on the management of the collapsed tri-athlete

I, _____, agree voluntarily to participate in the UCT/MRC Research Unit for Exercise Science and Sports Medicine's research project titled "**A study on the management of the collapsed tri-athlete**" performed by the University of Cape Town and the Sports Science Institute of South Africa. The following procedures and concepts have been explained to me in full:

1. Completion of a questionnaire: The completion of a questionnaire is not associated with any risk. Questionnaire and other clinical data (paper and electronic) will be kept confidential, will be kept secure, and will not be made available to any party other than the research team without the consent of the individual subjects.
2. Blood sample collection at registration and after the race: The potential risks to subjects of blood collection are minimal and are related to 1) blood sample collection technique, and 2) the volume of blood collected prior to racing and the potential risk of a decreased performance in the race. The potential risks associated with blood collection technique from the ante-cubital veins are: infection, delayed healing, haematoma, physical pain, mental discomfort and injury to a nerve or a vessel. These risks are small and will be minimized by the use of trained phlebotomists, use of sterile techniques and the use of disposable, single use materials. The risk of decreased performance as a result of blood collection will be reduced by not subjecting any participant to the collection of a blood volume exceeding 15ml prior to the race.
3. Measurement of body weight before and after the race: Body weight will be measured using a standard electronic scale, and there is no risk associated with this procedure.
4. Treatment if I collapse after the race: If I collapse during or after the race I might receive either IV or oral fluids ad libitum (as much as I want) but according to my post-race blood sodium level. Optimum care will be provided to me according to the current standard of care. Treatment will cease when my laboratory values have returned to normal and I am alert and oriented. I will be transported to the local hospital if my condition requires more urgent medical attention.

I have read the preceding subject information sheet and understand the testing procedures outlined therein. I understand any accompanying risks and discomforts. Knowing these risks and discomforts and having had the opportunity to pose questions answered to my satisfaction, I hereby consent to participate in this study. I understand that I may withdraw from this study at any time without further question. I have been informed that the individual

data derived from my participation in these protocols will remain confidential. I understand that the medical staff and the research team have professional medical insurance.

Name of the tri-athlete: _____

Signature of tri-athlete _____

Date: _____

Name of investigator: _____

Signature of Investigator: _____

Date: _____ _____

INFORMED CONSENT FORM

Component 2: A study to determine the cause of Exercise Associated Muscle Cramping (EAMC)

I, _____, agree voluntarily to participate in the UCT/MRC Research Unit for Exercise Science and Sports Medicine's research project titled "**A study to determine the cause of Exercise Associated Muscle Cramping (EAMC)**" performed by the University of Cape Town and the Sports Science Institute of South Africa. The following procedures and concepts have been explained to me in full:

1. Completion of a questionnaire: The completion of a questionnaire is not associated with any risk. Questionnaire and other clinical data (paper and electronic) will be kept confidential, will be kept secure, and will not be made available to any party other than the research team without the consent of the individual subjects.
2. Blood sample collection at registration and after the race: The potential risks to subjects of blood collection are minimal and are related to 1) blood sample collection technique, and 2) the volume of blood collected prior to racing and the potential risk of a decreased performance in the race. The potential risks associated with blood collection technique from the ante-cubital veins are: infection, delayed healing, haematoma, physical pain, mental discomfort and injury to a nerve or a vessel. These risks are small and will be minimized by the use of trained phlebotomists, use of sterile techniques and the use of disposable, single use materials. The risk of decreased performance as a result of blood collection will be reduced by not subjecting any participant to the collection of a blood volume exceeding 15ml prior to the race.
3. Measurement of body weight before and after the race: Body weight will be measured using a standard electronic scale, and there is no risk associated with this procedure.
4. Treatment if I develop muscle cramps during or after the race: If I develop muscle cramps during after the race I will receive treatment in a designated area of the medical facility. Optimum care will be provided to me according to the current standard of care. I will be required to have a rectal temperature measurement taken, blood samples will be collected, body weight will be measured, and I will have surface electrodes attached to my muscle to measure electrical activity. Treatment will cease when my cramps have stopped and I am able to stand up and walk. I will be transported to the local hospital if my condition requires more urgent medical attention.

I have read the preceding subject information sheet and understand the testing procedures outlined therein. I understand any accompanying risks and discomforts. Knowing these risks and discomforts and having had the opportunity to pose questions answered to my satisfaction, I hereby consent to participate in this study. I understand that I may withdraw from this study at any time without further question. I have been informed that the individual data derived from my participation in these protocols will remain confidential. I understand that the medical staff and the research team have professional medical insurance.

Name of the tri-athlete: _____

Signature of tri-athlete _____

Date: _____

Name of investigator: _____

Signature of Investigator: _____

Date: _____

INFORMED CONSENT FORM

Component 4: A study to determine the genetic basis for performance and physiological responses during an Ironman Triathlon

I, _____, agree voluntarily to participate in the UCT/MRC Research Unit for Exercise Science and Sports Medicine's research project titled "**A study to determine the genetic basis for performance and physiological responses during an Ironman Triathlon**" performed by the University of Cape Town and the Sports Science Institute of South Africa. The following procedures and concepts have been explained to me in full:

1. Completion of a questionnaire: The completion of a questionnaire is not associated with any risk. Questionnaire and other clinical data (paper and electronic) will be kept confidential, will be kept secure, and will not be made available to any party other than the research team without the consent of the individual subjects.
2. Blood sample collection at registration, and immediately after the race: The potential risks to subjects of blood collection are minimal and are related to 1) blood sample collection technique, and 2) the volume of blood collected prior to racing and the potential risk of a decreased performance in the race. The potential risks associated with blood collection technique from the ante-cubital veins are: infection, delayed healing, haematoma, physical pain, mental discomfort and injury to a nerve or a vessel. These risks are small and will be minimized by the use of trained phlebotomists, use of sterile techniques and the use of disposable, single use materials. The risk of decreased performance as a result of blood collection will be reduced by not subjecting any participant to the collection of a blood volume exceeding 15ml prior to the race.

I have read the preceding subject information sheet and understand the testing procedures outlined therein. I understand any accompanying risks and discomforts. Knowing these risks and discomforts and having had the opportunity to pose questions answered to my satisfaction, I hereby consent to participate in this study. I understand that I may withdraw from this study at any time without further question. I have been informed that the individual data derived from my participation in these protocols will remain confidential. I understand that the medical staff and the research team have professional medical insurance.

Name of the tri-athlete: _____

Signature of tri-athlete _____

Date: _____

Name of investigator: _____

Signature of Investigator: _____

Date: _____

INFORMED CONSENT FORM

Component 5: A study to determine the genetic risk/s associated with chronic Achilles tendon injuries in tri-athletes

I, _____, agree voluntarily to participate in the UCT/MRC Research Unit for Exercise Science and Sports Medicine's research project titled "**A study to determine the genetic risk/s associated with chronic Achilles tendon injuries in tri-athletes**" performed by the University of Cape Town and the Sports Science Institute of South Africa. The following procedures and concepts have been explained to me in full:

1. Completion of a questionnaire: The completion of a questionnaire is not associated with any risk. Questionnaire and other clinical data (paper and electronic) will be kept confidential, will be kept secure, and will not be made available to any party other than the research team without the consent of the individual subjects.
2. Blood sample collection at registration, and at the end of the race: The potential risks to subjects of blood collection are minimal and are related to 1) blood sample collection technique, and 2) the volume of blood collected prior to racing and the potential risk of a decreased performance in the race. The potential risks associated with blood collection technique from the ante-cubital veins are: infection, delayed healing, haematoma, physical pain, mental discomfort and injury to a nerve or a vessel. These risks are small and will be minimized by the use of trained phlebotomists, use of sterile techniques and the use of disposable, single use materials. The risk of decreased performance as a result of blood collection will be reduced by not subjecting any participant to the collection of a blood volume exceeding 15ml prior to the race.
3. Soft tissue diagnostic ultrasound examination: I understand that I will be subjected to a soft tissue diagnostic ultrasound examination of my Achilles tendons during the registration period, and on completion of the race. I understand that these investigations are not associated with any risk, and will be performed by a trained radiologist.

I have read the preceding subject information sheet and understand the testing procedures outlined therein. I understand any accompanying risks and discomforts. Knowing these risks and discomforts and having had the opportunity to pose questions answered to my satisfaction, I hereby consent to participate in this study. I understand that I may withdraw from this study at any time without further question. I have been informed that the individual data derived from my participation in these protocols will remain confidential. I understand that the medical staff and the research team have professional medical insurance.

Name of the tri-athlete: _____

Signature of tri-athlete _____

Date: _____

Name of investigator: _____

Signature of Investigator: _____

Date: _____

C02/200

INFORMED CONSENT FORM

I, _____, agree voluntarily to participate in the following components **(DELETE THOSE COMPONENTS YOU DO NOT AGREE TO PARTICIPATE IN)** of the UCT/MRC Research Unit for Exercise Science and Sports Medicine's, University of Cape Town, research project titled:-

1. "A study on the management of the collapsed tri-athlete",
2. "A study to determine the cause of Exercise Associated Muscle Cramping (EAMC)"
3. "A study on the management of pain in triathlon athletes",
4. "A study to determine the genetic basis for performance, physiological responses and medical complications during an Ironman Triathlon"
5. "A study to determine the extent of neural fatigue in athletes immediately post Ironman triathlon"
6. "Factors associated with gastro-intestinal (GIT) distress in Ironman triathletes"

I understand that my participation in this research project has no direct benefits to me during the Ironman 2007 competition. However, I understand that my participation in the research project will advance the medical and scientific knowledge related to endurance sports. Therefore, information gathered through my participation in this project could advance the future medical care, training advice and performance of endurance athletes.

I have read the subject information sheets and the following procedures and concepts have been explained to me in full:
(DELETE THOSE COMPONENTS YOU DO NOT AGREE TO PARTICIPATE IN)

1. **Completion of a questionnaire:** **(all components)**
   The completion of personal details, racing, training, equipment use, medical, supplement use, fluid use and lifestyle history questionnaires are not associated with any risk. Completion of self-rated behavioural questionnaires has not previously been shown to be associated with risk. A potential risk is that people who have experienced significant past trauma will find questionnaires on this uncomfortable. The questions within the behavioural questionnaires are asking are about temperament and none of the scales are directed at picking up psychopathology. Any personal identification of subjects (names and surnames), questionnaire data and other clinical data (paper and electronic) will be kept confidential, will be kept secure, and will not be made available to any party other than the research team without the consent of the individual subjects.

   I agree that the all the questionnaire information, my performance during the Ironman triathlon, together with all the other data collected from the various components of this trial may be used to answer scientific questions about the medical conditions,

physiological responses and measures of performance associated with the participation in and completion of an Ironman triathlon.

2. **Treatment if I collapse after the race: (only for the collapsed athlete component)**
   If I collapse during or after the race I might receive either fluid replacement directly into your vein or oral fluids ad libitum (as much as I want) but according to my post-race blood sodium level. Optimum care will be provided to me according to the current standard of care. Treatment will cease when my laboratory values have returned to normal and I am alert and oriented. I will be transported to the local hospital if my condition requires more urgent medical attention.

3. **Pre- and post-race serum electrolyte (salt) levels and weights (only for the collapsed athlete component)**
   I have agreed to donate 5 milliliters (1 teaspoons) of venous blood during registration and immediately after completing the race in the medical facility. The sample will be used to measure my serum electrolyte (blood salt) levels. The potential risks to subjects of blood collection ar I have agreed to donate ten milliliters (2 teaspoons) of venous blood. The sample will be used for the extraction and analysis of genetic material (DNA). mal and are related to 1) blood sample collection technique, and 2) the volume of blood collected prior to racing and the potential risk of a decreased performance in the race. The potential risks associated with blood collection technique from the ante-cubital veins are: infection, delayed healing, haematoma, physical pain, mental discomfort and injury to a nerve or a vessel. These risks are small and will be minimized by the use of trained phlebotomists, use of sterile techniques and the use of disposable, single use materials. The risk of decreased performance as a result of blood collection will be reduced by not subjecting any participant to the collection of a blood volume exceeding 15 ml prior to the race.

   Body weight will be measured on the morning before the start of the race and immediately after completing the race in the medical facility using a standard electronic scale, and there is no risk associated with this procedure.

4. **Measurement of heart rate data: (only for the cramps component)**
   This will be done with the subjects own heart rate monitor used during training and racing. The stored files will be emailed to the researcher at the Sports Science Institute, and will be kept confidential.

5. **<u>Score of perceived exertion during the race:</u> (only for the cramps and the management of pain components)**

During the race researchers will be allocated to about 12 stages throughout the race. As you swim, run or cycle past these researchers they will hold up two boards with the scores for *"perception of effort rating"*. You will be asked to shout out your respective scores as you go past them and they will record these scores against your race number. Data for this component of the study will involve contact with subjects during the race. There is a potential risk that in the process of data collection, the performance of subjects in the race will interfere with. This risk will be minimal, as the nature of the data collection is such that subjects will only be asked to shout out two numbers as they pass members of the research team at designated points in the race. However, should tri-athletes feel that this affects their performance during the race; they will be free to withdraw from this component of the study during the race. There will be no interference with other race participants during this data collection process.

6. **<u>Pain during the race:</u> (only for the management of pain components)**

During the race researchers will be allocated to about 12 stages throughout the race. As you swim, run or cycle past these researchers they will hold up two boards with the scores for *"pain assessment"*. You will be asked to shout out your respective scores as you go past them and they will record these scores against your race number. Data for this component of the study will involve contact with subjects during the race. There is a potential risk that in the process of data collection, the performance of subjects in the race will interfere with. This risk will be minimal, as the nature of the data collection is such that subjects will only be asked to shout out two numbers as they pass members of the research team at designated points in the race. However, should tri-athletes feel that this affects their performance during the race; they will be free to withdraw from this component of the study during the race. There will be no interference with other race participants during this data collection process.

7. **<u>Recording of heart rate variability during stroop test:</u> (only for the management of pain components)**

The stroop test is a simple, computer based test. The mental concentration that is required for the test is relevant for the data collection and not the outcome of the test. There is no risk associated with the recording of the heart rate variability

8. **<u>Pain threshold with a digital pain probe:</u> (only for the management of pain components)**

There is no risk associated with the assessment of the pain threshold with the digital pain probe. As the onset of pain is determined, the discomfort is minimal.

9. **Brain wave measurements: (only for the neural fatigue)**

There are no potential risks associated with brain wave measurements, since we are merely recording the underlying electric activity generated by the brain and not stimulating the brain in any way. Similarly, there are also no potential risks associated with measuring the electrical activity generated by the heart. There may be some discomfort experienced by the EEG gel needed to increase the conductivity of the electric signal, but no more so than what would be experienced by applyingbhair gel to flatten your hair.

10. **Blood sample collection for genetic studies: (only for the genetics component)**

At one of the pre-race facilities or at race registration, I have agreed to donate ten milliliters (2 teaspoons) of venous blood. The sample will be used for the extraction and analysis of genetic material (DNA).

The potential risks to subjects of blood collection are minimal and are related to 1) blood sample collection technique, and 2) the volume of blood collected prior to racing and the potential risk of a decreased performance in the race. The potential risks associated with blood collection technique from the ante-cubital veins are: infection, delayed healing, haematoma, physical pain, mental discomfort and injury to a nerve or a vessel. These risks are small and will be minimized by the use of trained phlebotomists, use of sterile techniques and the use of disposable, single use materials. The risk of decreased performance as a result of blood collection will be reduced by not subjecting any participant to the collection of a blood volume exceeding 15 ml prior to the race.

The DNA will only be used for scientific research purposes relating to the genetic basis of (1) athletic ability, (2) physiological response to (3) medical complications during ultra-endurance events. I have also agreed to complete personal particulars, training, sporting, measures of behavioural endophenotypes and medical questionnaires and understand that all the information that is collected during the study will be treated with the strictest confidentiality and will only be used for scientific research purposes. Questionnaire and other clinical data (paper and electronic) will be kept confidential, will be kept secure, and will not be made available to any party other than the research team without the consent of the individual subjects. I also understand that all data will be analysed anonymously and my DNA sample will be destroyed on completion of the study.

I understand that some of the DNA extracted form the donated blood sample will be sent to the Cyprus Institute of Neurology and Genetics in Cyprus for analysis. I understand that the DNA samples will be shipped to and analysed in Cyprus anonymously. I understand that the DNA will be genotyped (analysed) for variations (polymorphisms)

within genes relating to the genetic basis of athletic ability, tendon and ligament overuse injuries and dysnatraemia during ultra-endurance events only.

I understand that whilst there is no direct benefit to myself, if a genetic predisposition for (1) athletic ability, (2) physiological response to and (3) medical complications during ultra-endurance events can be established, then future generations will be able to establish their risk for this condition. This may allow better prevention and treatment options in the future. I understand that I will receive the overall results of the study.

I have read (or, where appropriate, have had read to me) and understood the information about this study, and any questions I have asked have been answered to my satisfaction. I agree to participate in the study, realising that I have the right to request that my DNA sample be destroyed at anytime. I agree that research data provided by me or with my permission during the project may be included in a thesis, presented at conferences and published in journals on the condition that neither my name nor any other identifying information is used.

11. **Abdominal ultrasound to determine blood flow to the abdominal organs (only for the GIT component)**

I agree to having a pre-and post-race abdominal ultrasound to measure the blood flow to my abdominal organs.

I have read the preceding subject information sheet and understand the testing procedures outlined therein. I understand any accompanying risks and discomforts. Knowing these risks and discomforts and having had the opportunity to pose questions answered to my satisfaction, I hereby consent to participate in this study. I understand that I may withdraw from this study at any time without further question. I have been informed that the individual data derived from my participation in these protocols will remain confidential. I understand that the medical staff and the research team have professional medical insurance.

Name of the tri-athlete: _____

Signature of tri-athlete _____

Date: _____

Name of investigator: _____Prof Martin Schwellnus____

Signature of Investigator: _____

Date: _____ _____

4. Participant questionnaires

Questionnaire used for the data collection for Chapter 3. Anterior Cruciate Ligament Ruptures.

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: +27 21 650 4561
Fax: +27 21 686 7530

GENETIC BASIS OF LIGAMENT INJURY QUESTIONNAIRES

A. PERSONAL PARTICULARS			
Surname			
First Name			
Postal Address		Code	
E-mail address		Phone (day time)	
Date of birth	Y Y Y Y / M M / D D	Cell	
Height (cm)		Gender	Male ☐ Female ☐
Weight (kg)	Pre-Injury:	Current:	
Ethnic group (Only Required and Used for Research Purposes)	Black/African ☐ White ☐ Indian ☐ Mixed Ancestry (Coloured) ☐ Asian ☐ Other ☐		
Ancestry: Tribal or national background (eg Xhosa, Dutch, Zulu, German, Italian)	Father Unknown ☐		
	Mother Unknown ☐		
Country of Birth			
Dominant Hand	Left ☐ Right ☐ Ambi ☐	Dominant Leg	Left ☐ Right ☐ Ambi ☐
Smoker	Yes (Current) ☐	Yes (Ex smoker) ☐	No, never ☐
	If yes, Number of years _____	If stopped, when _____	
	If yes, number per day _____		

(If you participate or have participated in more than 6 sports, please complete additional Sporting Details Questionnaires, Part B)

B. SPORTING DETAILS Please record your sporting activities in order of importance			
Type of sport(s) you have participated in (please name)	**Main sport 1**	**Other sport 2**	**Other sport 3**
Current or past participation	Current ☐ Past ☐	Current ☐ Past ☐	Current ☐ Past ☐
Year started participation			
Number of years involved in the sport			
Position played prior to injury (if appropriate)			
Playing level prior to injury (if appropriate)			
Number of years played prior to the injury.			

Type of sport(s) you have participated in (please name)	**Other sport 4**	**Other sport 5**	**Other sport 6**
Current or past participation	Current ☐ Past ☐	Current ☐ Past ☐	Current ☐ Past ☐
Year started participation			
Number of years involved in the sport			
Position played prior to injury (if appropriate)			
Playing level prior to injury (if appropriate)			
Number of years played prior to the injury.			

C. ANTERIOR CRUTIATE LIGAMENT INJURY DETAILS

Date of ACL injury?	
Which side was injured?	☐ Left ☐ Right ☐ Both
To what extent was your ligament ruptured?	☐ Complete ☐ Partial ☐ None ☐ Unknown
Investigation done to confirm the diagnosis	☐ MRI ☐ Surgery
How bad is your pain today? (mark line: e.g.|----------I----------|)	|----------| No pain — Pain as bad as it can be
How was the ACL ruptured? **(please also explain exactly how the injury occurred)**	☐ Direct impact ☐ Twisting and bending with contact ☐ Twisting and bending without contact ☐ Other non-contact ☐ Other……
What was the initial treatment? (You may tick more than one block.)	☐ Ice application ☐ Compresion ☐ Immobilisation ☐ Medication ☐ Other……
What was the final treatment?	☐ Surgery ☐ Reabilitation ☐ Other……
What are your current symptoms? (You may tick more than one block.)	☐ Pain ☐ Swelling ☐ Instability ☐ Weakness ☐ Other……
What is your current sports participation?	☐ None ☐ Limited to non-weight bearing exercise ☐ Limited, not to same level as pre-injury ☐ Full participation

If you are able to recall, what were the weather and pitch conditions like at the time of injury?	☐ Wet and soft ground ☐ Dry, but soft ground ☐ Dry and firm ground ☐ Wet, but firm ground ☐ Other……
Associated injuries?	☐ Meniscal tear ☐ MCL tear ☐ Other ligament tear ☐ Bone bruising ☐ Other……

D. HISTORY OF OTHER LIGAMENT AND TENDON INJURIES IN THE PAST				
Have you ever injured a ligament in the past?	Yes ☐ No ☐			
If yes, please specify which ligaments? (You may tick more than one block, please select either L (left) or R (right))		L R		L R
	Knee (ACL)	☐ ☐	Wrist ligaments	☐ ☐
	Knee (MCL)	☐ ☐	Finger ligaments	☐ ☐
	Ankle lateral ligaments	☐ ☐	Knee (PCL)	☐ ☐
	Spinal ligaments	☐ ☐	Knee (LCL)	☐ ☐
	Shoulder ligaments	☐ ☐	Ankle medial ligaments	☐ ☐
	Elbow ligaments	☐ ☐	Other ligaments	☐ ☐
To your knowledge, have any other members of your family suffered from any ligament injury?	Yes ☐ No ☐	If Yes, please specify the family member ☐ Mother ☐ Father ☐ Sibling ☐ Son / daughter ☐ Other family member…… and condition: Please choose ligament injury from the list above ……		
Have you ever injured a tendon in the past?	Yes ☐ No ☐			

<table>
<tr><td rowspan="9">If yes, please specify which tendon?
(You may tick more than one block, please select either L (left) or R (right))</td><td rowspan="3">Foot and ankle:</td><td></td><td>L</td><td>R</td></tr>
<tr><td>Achilles tendon</td><td>☐</td><td>☐</td></tr>
<tr><td>Tibialis posterior</td><td>☐</td><td>☐</td></tr>
<tr><td>Plantar fascia</td><td>☐</td><td>☐</td></tr>
<tr><td>Knee:</td><td>Patellar tendon</td><td>☐</td><td>☐</td></tr>
<tr><td>Elbow and wrist:</td><td>Wrist extensor tendons</td><td>☐</td><td>☐</td></tr>
<tr><td rowspan="4">Shoulder:</td><td>Subscapularis</td><td>☐</td><td>☐</td></tr>
<tr><td>Supraspinatus</td><td>☐</td><td>☐</td></tr>
<tr><td>Infraspinatus</td><td>☐</td><td>☐</td></tr>
<tr><td>Teres minor</td><td>☐</td><td>☐</td></tr>
<tr><td colspan="4">Other:.....</td></tr>
<tr><td>To your knowledge, have any other members of your family suffered from any tendon pathology?</td><td colspan="4">Yes ☐ No ☐
If Yes, please specify the family member
☐ Mother
☐ Father
☐ Sibling
☐ Son / daughter
☐ Other family member:.....
Condition: Please choose tendon injury from the list above
.....</td></tr>
<tr><td>Have you ever suffered from any of the following joint capsule injuries?</td><td colspan="4">☐ Acute shoulder dislocation
☐ Chronic shoulder instability
☐ Chronic ankle instability
☐ Other: _____
_____</td></tr>
</table>

E. MEDICAL HISTORY

Do you currently suffer from any of these medical conditions:

☐ High Blood Pressure	☐ Angina/Heart Attack	☐ Asthma
☐ Emphysema	☐ Rheumatoid arthritis	☐ Osteoarthritis (wear & tear)
☐ Malignant disease (cancer)	☐ Elevated Blood Cholesterol	☐ Adrenal disorders
	☐ Diabetes mellitus	☐ Thyroid disorders
If Yes, what type?	☐ Renal disease	☐ Amyloidosis
_____		

Do you currently suffer from any other Connective Tissue & Rheumatological Diseases & Disorders?	Yes ☐ No ☐	If Yes, please select from the list below

List of some Connective Tissue and/or Rheumatic Diseases and Disorders

☐ Ankylosing Spondylitis	☐ Lipid Storage Diseases	☐ Pseudogout
☐ Aspartylglycosaminuria (AGU)	☐ Marfan Syndrome	☐ Reactive Arthritis
☐ Behcet's Syndrome	☐ Menkes Kinky Hair Syndrome	☐ Reiter's Syndrome
☐ Crohn's Disease	☐ Mucopolysaccharidoses	☐ Relapsing Polychrondritis
☐ Discoid Lupus Erythematosus	☐ Myopathies and Dystrophies	☐ Scleroderma
☐ Ehlers-Danlos syndrome (EDS)	☐ Ochronosis (Homocystinuria)	☐ Sjogren's Syndrome
☐ Eosinophilic Fascitis	☐ Osteogenesis imperfecta (OI)	☐ Systemic Lupus Erythematosus (SLE)
☐ Giant Cell (Temporal) Arthritis	☐ Polyarteritis Nodosa	☐ Systemic Sclerosis
☐ Gout	☐ Polymyalgia Rheumatica	☐ Wegener's Granulomatosis
☐ Hypersentive Vasulatis	☐ Polymyositis & Dermatomyositis	☐ Other_____

What surgical operations have you had? (please list and give dates)	Operation	Date

If female:

At what age did you start menstruating? (years)	
Are you currently using any type of contraception?	☐ Yes ☐ No
If Yes, what type of contraception are you using?	☐ Pill ☐ Injection ☐ IUD

Are you currently?	☐ Pre-menopausal (±12 cycles per year at intervals of 23–33 days & bleeding lasts 3-7 days) ☐ Menopausal (cycles are irregular and less frequent) ☐ Post-menopausal (no longer menstrusting)	
Family History		
Do any other members of your family suffer from elevated blood cholesterol?	Yes ☐ No ☐	If Yes, which relative? ☐ Mother ☐ Father ☐ Sibling ☐ Son / daughter ☐ Other relative;.....
Is there any history of arthritis in your family?	Yes ☐ No ☐	If Yes, which relative? ☐ Mother ☐ Father ☐ Sibling ☐ Son / daughter ☐ Other relative;..... & What type of aethritis? Rheumatoid ☐ Osteoarthritis ☐ Other ☐

Drug and Allergy History	If yes, how long ago (or how many times, where applicable) did you use the medication?	
Have you ever used oral corticosteroids (cortisone tablets)?	Yes ☐ No ☐	☐ 3 months ☐ 6 months ☐ 12 months ☐ 24 or more months
Have you ever been given an injection with corticosteroids?	Yes ☐ No ☐	☐ 3 months ☐ 6 months ☐ 12 months ☐ 24 or more months
Have you ever been given an injection of corticosteroids in or around a tendon?	Yes ☐ No ☐	☐ Once ☐ Twice ☐ 3 times ☐ >3 times
Have you ever used anabolic steroids?	Yes ☐ No ☐	☐ 3 months ☐ 6 months ☐ 12 months ☐ 24 or more months
Have you ever used fluoroquinoline antibiotics?	Yes ☐ No ☐	☐ 3 months ☐ 6 months ☐ 12 months ☐ 24 or more months

If yes, please select from the list below:		
☐ ADCO-CIPRIN	☐ CIPROBAY	☐ SANDOZ CIPROFLOXACIN
☐ AVELON	☐ CIPROGEN	☐ TAFLOC
☐ BACTIDRON	☐ CPL ALLIANCE CIPROFLOXACIN	☐ TARIVID
☐ CIFLOC	☐ DYNAFLOC	☐ TAVANIC
☐ CIFRAN	☐ FLOXIN	☐ TEQUIN
☐ CIPLA-CIPROFLOXACIN	☐ MAXAQUIN	☐ UNIQUIN
☐ CIPLOXX	☐ NOROXIN	☐ UTN-400
☐ CIPRO-HEXAL	☐ ORPIC	☐ ZANOCIN
☐ Other_____		

What medication, if any, are you currently using? (please list)	
What allergies do you have? (please list)	

F. OCCUPATIONAL DETAILS	
What is your current occupation?	
What was your occupation prior to injuring your ligament?	
Prior to injury, did your occupation involve lower limb activity?	☐ Yes ☐ No
If yes please indicate which legs.	Right leg ☐ Both legs ☐ Left leg ☐ None ☐

Questionnaire used for the data collection for Chapter 4. Achilles Tendinopathy

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: +27 21 650 4561
Fax: +27 21 686 7530

GENETIC BASIS OF TENDON INJURY QUESTIONNAIRES

A. PERSONAL PARTICULARS			
Surname			
First Name			
Postal Address		Code	
E-mail address		Phone (day time)	
Date of birth	Y Y Y Y / M M / D D	Cell	
Height (cm)		Gender	Male ☐ Female ☐
Weight (kg)	Pre-Injury:	Current:	
Ethnic group (Only Required and Used for Research Purposes)	Black/African ☐ White ☐ Indian ☐ Mixed Ancestry (Coloured) ☐ Asian ☐ Other ☐		
Ancestry: Tribal or national background	Father	Unknown ☐	
	Mother	Unknown ☐	
Country of Birth			
Dominant Hand	Left ☐ Right ☐ Ambi ☐	Dominant Leg	Left ☐ Right ☐ Ambi ☐
Smoker	Yes (Current) ☐	Yes (Ex smoker) ☐	No, never ☐
	If yes, Number of years _____	If stopped, when _____	
	If yes, number per day _____		

(If you participate or have participated in more than 6 sports, please complete additional Sporting Details Questionnaires, Part B)

B. SPORTING DETAILS Please record your sporting activities in order of importance			
Type of sport(s) you have participated in (please name)	**Main sport 1**	**Other sport 2**	**Other sport 3**
Current or past participation	Current ☐ Past ☐	Current ☐ Past ☐	Current ☐ Past ☐
Year started participation			
Number of years involved in the sport			
Position played prior to injury (if appropriate)			
Playing level prior to injury (if appropriate)			
Number of years played prior to the injury.			

Type of sport(s) you have participated in (please name)	**Other sport 4**	**Other sport 5**	**Other sport 6**
Current or past participation	Current ☐ Past ☐	Current ☐ Past ☐	Current ☐ Past ☐
Year started participation			
Number of years involved in the sport			
Position played prior to injury (if appropriate)			
Playing level prior to injury (if appropriate)			
Number of years played prior to the injury.			

C. FLEXIBILITY TRAINING HISTORY	
Do you perform flexibility training (regular stretching exercises)?	Yes ☐ No ☐
If YES, please complete the rest of the flexibility training history section below:- If NO, continue completing the questionnaire from the top of page 5 (Equipment use history).	
On average, how many days a week do you perform a stretching session?	days/week

On average, how times a day do you perform a stretching session?	times/day
Please tick which muscle groups do you include in your stretching session?	☐ Hamstrings ☐ Quadriceps ☐ Calf (gastrocnemius) ☐ Calf (soleus) ☐ Groin (inner thigh) ☐ Upper body limbs ☐ Other: _____
Please tick when you stretch? (before, during and/or after exercising. You can tick more than one box)	☐ Before Exercise ☐ During Exercise ☐ After Exercise
When you stretch an individual muscle group, on average, **how long do you hold the stretch** for?	seconds
When you stretch an individual muscle group, on average, **how many times do you stretch the muscle for**?	☐ Once ☐ Twice ☐ 3 times ☐ 4 times ☐ 5 times ☐ 6 or more times

D. TENDON INJURY - MEDICAL DETAILS

Symptoms

How many times have you had tendon injuries?	Tendon Injured	Date of Injury	Acute or Chronic Injury	Sudden[1] or Gradual[2] Onset
[1]Sudden onset is within a few seconds or minutes	1			
	2			
[2]Gradual onset is over days or weeks	3			
	4			
	5			

Please complete a **separate form**, Part D only, for each Tendon Injury you have had	
Injury Number (1,2,3,4,or 5)	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ _____
Which tendon did you injure?	Rotator cuff tendon ☐ • Supraspinatus ☐ • Infraspinatus ☐ • teres minor ☐ Patellar tendon ☐ Wrist extensor tendons ☐ Achilles tendon ☐
Which side was injured?	☐ Left ☐ Right ☐ Both
Which region of your tendon was injured? Please indicate on a diagram. (Only if applicable)	☐ Upper 1/3 ☐ Middle 1/3 ☐ Lower 1/3

To what extent was your Tendon ruptured?	☐ Complete ☐ Partial ☐ None
How were you injured? (e.g. sport, walking)	
Grade of injury **at the time** of injury	☐ pain only after exercise ☐ pain during exercise, but did not cause you to alter training ☐ pain during exercise, which causes you to alter training ☐ pain which causes you to stop training ☐ no pain ☐ not sure ☐ Other (Specify _____)
Grade of injury **currently**	☐ pain only after exercise ☐ pain during exercise, but did not cause you to alter training. ☐ pain during exercise, which causes you to alter training ☐ pain which causes you to stop training ☐ no pain ☐ not sure ☐ Other (Specify _____)
Which of the following symptoms were present **before** the injury	☐ Pain (less than 1 week) ☐ Stiffness ☐ Pain (1-4 weeks) ☐ Swelling ☐ Pain (> 4 weeks) ☐ None
Which of the following symptoms were present **after** the injury	☐ Pain (less than 1 week) ☐ Stiffness ☐ Pain (1-4 weeks) ☐ Swelling ☐ Pain (> 4 weeks) ☐ None
If you have or had chronic tendon pain, what seems to alleviate the pain?	

Diagnosis	
Which type of Tendon Disease were you diagnosed with e.g. Rupture, Tendinitis, etc.	
Diagnosed by (Please indicate the name and contact number of the clinician who diagnosed you)	☐ Doctor _____ ☐ Physiotherapist _____ ☐ Biokineticist _____ ☐ Podiatrist _____ ☐ Other _____
If you had a tendon rupture. How was it treated?	☐ Surgically ☐ Non-surgically
If applicable, who was the surgeon?	Surgeon _____ Phone _____
If applicable, what diagnostic imaging was performed?	☐ Ultrasound ☐ MRI ☐ CT Other _____
If applicable, who did the imaging?	Clinician _____ Phone _____

E. HISTORY OF OTHER LIGAMENT AND TENDON INJURIES IN THE PAST					
Have you ever injured a ligament in the past?	Yes ☐ No ☐				
If yes, please specify which ligaments? (You may tick more than one block, please select either L (left) or R (right))		L	R		L R
	Knee (ACL)	☐	☐	Wrist ligaments	☐ ☐
	Knee (MCL)	☐	☐	Finger ligaments	☐ ☐
	Ankle lateral ligaments	☐	☐	Knee (PCL)	☐ ☐
	Spinal ligaments	☐	☐	Knee (LCL)	☐ ☐
	Shoulder ligaments	☐	☐	Ankle medial ligaments	☐ ☐
	Elbow ligaments	☐	☐	Other ligaments	☐ ☐
To your knowledge, have any other members of your family suffered from any ligament injury?	Yes ☐ No ☐	If Yes, please specify the family member ☐ Mother ☐ Father ☐ Sibling ☐ Son / daughter ☐ Other family member…… and condition: Please choose ligament injury from the list above ……			
Have you ever injured a tendon in the past?	Yes ☐ No ☐				
If yes, please specify which tendon? (You may tick more than one block, please select either L (left) or R (right))			L R		
	Foot and ankle:	Achilles tendon	☐ ☐		
		Tibialis posterior	☐ ☐		
		Plantar fascia	☐ ☐		
	Knee:	Patellar tendon	☐ ☐		
	Elbow and wrist:	Wrist extensor tendons	☐ ☐		
	Shoulder:	Subscapularis	☐ ☐		
		Supraspinatus	☐ ☐		
		Infraspinatus	☐ ☐		
		Teres minor	☐ ☐		
	Other:……				

To your knowledge, have any other members of your family suffered from any tendon pathology?	Yes ☐ No ☐	If Yes, please specify the family member ☐ Mother ☐ Father ☐ Sibling ☐ Son / daughter ☐ Other family member:..... Condition: Please choose tendon injury from the list above .....
Have you ever suffered from any of the following joint capsule injuries?	☐ Acute shoulder dislocation ☐ Chronic shoulder instability ☐ Chronic ankle instability ☐ Other: _____ _____	

F. MEDICAL HISTORY

Do you currently suffer from any of these medical conditions:

☐ High Blood Pressure	☐ Angina/Heart Attack	☐ Asthma
☐ Emphysema	☐ Rheumatoid arthritis	☐ Osteoarthritis (wear & tear)
☐ Malignant disease (cancer)	☐ Elevated Blood Cholesterol	☐ Adrenal disorders
If Yes, what type?	☐ Diabetes mellitus	☐ Thyroid disorders
_____	☐ Renal disease	☐ Amyloidosis

Do you currently suffer from any other Connective Tissue & Rheumatological Diseases & Disorders?	Yes ☐ No ☐	If Yes, please select from the list below

List of some Connective Tissue and/or Rheumatic Diseases and Disorders

☐ Ankylosing Spondylitis	☐ Lipid Storage Diseases	☐ Pseudogout
☐ Aspartylglycosaminuria (AGU)	☐ Marfan Syndrome	☐ Reactive Arthritis
☐ Behcet's Syndrome	☐ Menkes Kinky Hair Syndrome	☐ Reiter's Syndrome
☐ Crohn's Disease	☐ Mucopolysaccharidoses	☐ Relapsing Polychrondritis
☐ Discoid Lupus Erythematosus	☐ Myopathies and Dystrophies	☐ Scleroderma
☐ Ehlers-Danlos syndrome (EDS)	☐ Ochronosis (Homocystinuria)	☐ Sjogren's Syndrome
☐ Eosinophilic Fascitis	☐ Osteogenesis imperfecta (OI)	☐ Systemic Lupus Erythematosus (SLE)
☐ Giant Cell (Temporal) Arthritis	☐ Polyarteritis Nodosa	☐ Systemic Sclerosis
☐ Gout	☐ Polymyalgia Rheumatica	☐ Wegener's Granulomatosis
☐ Hypersentive Vasulatis	☐ Polymyositis & Dermatomyositis	☐ Other_____

What surgical operations have you had? (please list and give dates)	Operation	Date
If female:		
At what age did you start menstruating? (years)		
Are you currently using any type of contraception?	☐ Yes ☐ No	
If Yes, what type of contraception are you using?	☐ Pill ☐ Injection ☐ IUD	
Are you currently?	☐ Pre-menopausal (±12 cycles per year at intervals of 23–33 days & bleeding lasts 3-7 days) ☐ Menopausal (cycles are irregular and less frequent) ☐ Post-menopausal (no longer menstruating)	
Family History		
Do any other members of your family suffer from elevated blood cholesterol?	Yes ☐ No ☐	If Yes, which relative? ☐ Mother ☐ Father ☐ Sibling ☐ Son / daughter ☐ Other relative:.....
Is there any history of arthritis in your family?	Yes ☐ No ☐	If Yes, which relative? ☐ Mother ☐ Father ☐ Sibling ☐ Son / daughter ☐ Other relative:..... & What type of aethritis? Rheumatoid ☐ Osteoarthritis ☐ Other ☐

Drug and Allergy History	If yes, how long ago (or how many times, where applicable) did you use the medication?	
Have you ever used oral corticosteroids (cortisone tablets)?	Yes ☐ No ☐	☐ 3 months ☐ 6 months ☐ 12 months ☐ 24 or more months
Have you ever been given an injection with corticosteroids?	Yes ☐ No ☐	☐ 3 months ☐ 6 months ☐ 12 months ☐ 24 or more months
Have you ever been given an injection of corticosteroids in or around the Achilles tendon?	Yes ☐ No ☐	☐ Once ☐ Twice ☐ 3 times ☐ >3 times
Have you ever used anabolic steroids?	Yes ☐ No ☐	☐ 3 months ☐ 6 months ☐ 12 months ☐ 24 or more months
Have you ever used fluoroquinoline antibiotics?	Yes ☐ No ☐	☐ 3 months ☐ 6 months ☐ 12 months ☐ 24 or more months

If yes, please select from the list below:

☐ ADCO-CIPRIN	☐ CIPROBAY	☐ SANDOZ CIPROFLOXACIN
☐ AVELON	☐ CIPROGEN	☐ TAFLOC
☐ BACTIDRON	☐ CPL ALLIANCE CIPROFLOXACIN	☐ TARIVID
☐ CIFLOC	☐ DYNAFLOC	☐ TAVANIC
☐ CIFRAN	☐ FLOXIN	☐ TEQUIN
☐ CIPLA-CIPROFLOXACIN	☐ MAXAQUIN	☐ UNIQUIN
☐ CIPLOXX	☐ NOROXIN	☐ UTN-400
☐ CIPRO-HEXAL	☐ ORPIC	☐ ZANOCIN
☐ Other_____		

What medication, if any, are you currently using? (please list)	
What allergies do you have? (please list)	

G. OCCUPATIONAL DETAILS	
What is your current occupation?	
What was your occupation prior to injuring your tendon?	
Prior to injury, did your occupation involve lower limb activity?	☐ Yes ☐ No
If yes please indicate which legs.	Right leg ☐ Both legs ☐ Left leg ☐ None ☐

Questionnaire used for the data collection for Chapter 5. Range of Motion

GENETIC BASIS OF RANGE OF MOTION MEASUREMENTS QUESTIONAIRES

PERSONAL DETAILS			
Surname			
First Name			
Postal Address		Postal/ Zip Code	
E-mail address		Phone (day time)	code number
Alternate E-mail address			
Date of birth	yyyy-mm-dd	Cell (Mobile)	
Height	cm	Gender	Male ☐ Female ☐
Weight	kg	Age	yrs
Ethnic group (Only Required and Used for Research Purposes)	Black/African ☐ White ☐ Indian ☐ Mixed Ancestry (Coloured) ☐ Asian ☐ Other ☐		
Ancestry: Tribal or national background (eg Xhosa, Dutch, Zulu, German, Italian)	Father: Unknown ☐ Mother: Unknown ☐		
Country of Birth			
Dominant Hand	Left ☐ Right ☐ Both ☐	Dominant Leg	Left ☐ Right ☐ Both ☐
Occupation			
What **percentage** of your **working** day is spent in the following activities?	Sitting: ____ % Standing: ____ % Walking (Lower body activity) ____ % Manual Labour (upper and body activity) ____ %		

1

(If you participate or have participated in more than 6 sports, please complete additional Sporting Details Questionnaires)

SPORTING DETAILS Please record your sporting activities in order of importance			
Type of sport(s) you have participated in (please name)	**Main Sport 1**	**Other Sport 2**	**Other Sport 3**
Current or past participation	Current ☐ Past ☐	Current ☐ Past ☐	Current ☐ Past ☐
Year started participation			
Number of years involved in the sport			
Average hours of training per week			
Position played (if appropriate)			
Playing level (if appropriate)			

Type of sport(s) you have participated in (please name)	**Other Sport 4**	**Other Sport 5**	**Other Sport 6**
Current or past participation	Current ☐ Past ☐	Current ☐ Past ☐	Current ☐ Past ☐
Year started participation			
Number of years involved in the sport			
Average hours of training per week			
Position played (if appropriate)			
Playing level (if appropriate)			

FLEXIBILITY TRAINING HISTORY	
Do you perform flexibility training (regular stretching exercises)?	Yes ☐ No ☐
If YES, please complete the rest of the flexibility training history section below:- If NO, continue completing the questionnaire from the top of page 5 (Equipment use history)	
On average, how many days a week do you perform a stretching session?	days/week
On average, how many times a day do you perform a stretching session?	times/day
Please tick which muscle groups do you include in your stretching session?	☐ Hamstrings ☐ Quadriceps ☐ Calf (gastrocnemius) ☐ Calf (soleus) ☐ Groin (inner thigh) ☐ Upper body limbs ☐ Other: _____
Please tick when you stretch? (Before, during and/or after exercising. You can tick more than one box)	☐ Before Exercise ☐ During Exercise ☐ After Exercise
When you stretch an individual muscle group, on average, **how long do you hold the stretch** for?	seconds
When you stretch an individual muscle group, on average, **how many times do you stretch the muscle for**?	☐ Once ☐ Twice ☐ 3 times ☐ 4 times ☐ 5 times ☐ 6 or more times

PERSONAL GENERAL MEDICAL HISTORY	
1. Do you currently suffer from high blood pressure?	Yes ☐ No ☐
2. Do you currently suffer from angina/heart attack?	Yes ☐ No ☐
3. Do you currently suffer from asthma?	Yes ☐ No ☐
4. Do you currently suffer from emphysema?	Yes ☐ No ☐
5. Do you currently suffer from rheumatoid arthritis?	Yes ☐ No ☐
6. Do you currently suffer from osteoarthritis (wear and tear)?	Yes ☐ No ☐
7. Do you currently suffer from malignant disease (cancer)?	Yes ☐ No ☐
8. Do you currently suffer from elevated blood cholesterol?	Yes ☐ No ☐
9. Do you currently suffer from adrenal disorders?	Yes ☐ No ☐
10. Do you currently suffer from diabetes mellitus?	Yes ☐ No ☐
11. Do you currently suffer from thyroid disorders?	Yes ☐ No ☐
12. Do you currently suffer from renal disease?	Yes ☐ No ☐
13. Do you currently suffer from amyloidosis?	Yes ☐ No ☐

14. Please tick in which anatomical area you had **surgery** performed, if ever.	☐ Gastric (stomach)	☐ Oesophageal (swallowing pipe)
	☐ Small bowel	☐ Large bowel (colon)
	☐ Rectum	☐ Gallbladder
	☐ Pancreas	☐ Liver
	☐ Abdomen (general)	☐ Wrist
	☐ Head	☐ Finger
	☐ Neck	☐ Lower back
	☐ Face	☐ Hip
	☐ Front chest	☐ Thigh
	☐ Back chest	☐ Knee
	☐ Shoulder	☐ Lower leg
	☐ Upper arm	☐ Achilles
	☐ Elbow	☐ Ankle
	☐ Forearm	☐ Foot
	☐ Other (Specify: _____)	

History of medication and supplement use			
What medication, if any, are you currently using? (please list)	Name of medication		Years taken
Have you ever used oral corticosteroids (cortisone tablets)? (If **yes**, how long ago?)	Yes ☐ No ☐	☐ 3 months ☐ 12 months	☐ 6 months ☐ 24 or more months
Have you ever been given an injection with corticosteroids? (If **yes**, how long ago?)	Yes ☐ No ☐	☐ 3 months ☐ 12 months	☐ 6 months ☐ 24 or more months
Have you ever been given an injection of corticosteroids in or around the **Achilles** tendon? (If **yes**, how many times?)	Yes ☐ No ☐	☐ Once ☐ 3 times	☐ Twice ☐ >3 times
Have you ever used fluoroquinoline antibiotics? (refer to the following list)	Yes ☐ No ☐	☐ 3 months ☐ 12 months	☐ 6 months ☐ 24 or more months
What medication, if any, are you currently using? (please list)			
What allergies do you have? (please list)			

List of some fluoroquinoline antibiotics:		
ADCO-CIPRIN	CIPROBAY	SANDOZ CIPROFLOXACIN
AVELON	CIPROGEN	TAFLOC
BACTIDRON	CPL ALLIANCE CIPROFLOXACIN	TARIVID
CIFLOC	DYNAFLOC	TAVANIC
CIFRAN	FACTIVE	TEQUIN
CIPLA-CIPROFLOXACIN	FLOXIN	UNIQUIN
CIPLOXX	MAXAQUIN	UTIN-400
CIPRO-HEXAL	NOROXIN	ZANOCIN
	ORPIC	

LIFESTYLE AND HABITS HISTORY			
Please indicate your smoking status	Current smoker ☐	Ex smoker ☐	Never smoked ☐
If you answered yes, (past or current smoker) please complete the section on the right	Number of years of smoking:	If stopped, how many years ago:	
	What is (was) the average number of cigarettes per day:		

FAMILY MEDICAL HISTORY

Have any of your blood (biological) relatives ever had the following?

Please tick yes or no. If yes, please tick the relationship of that person to you (You may tick more than one of the relationship blocks).

Description		If Yes, please indicate the relationship
Chronic Achilles tendon injury	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Achilles tendon rupture	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Any ligament injury	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Arthritis	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Elevated blood cholesterol	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother

TENDON AND LIGAMENT INJURY HISTORY

Please tick which **tendon/s** you have injured? (next column on the tright)

Also indicate (tick) if your injured tendon was longsatnding pain (tendinopathy) or an acute tear/rupture

	Tendon	L	R	Longstanding Pain (Tendinopathy)	Acute Tear/ Rupture
Foot and ankle:	☐ Achilles tendon	☐	☐	☐	☐
	☐ Tibialis posterior	☐	☐	☐	☐
	☐ Plantar fascia	☐	☐	☐	☐
Knee:	☐ Patellar tendon	☐	☐	☐	☐
Elbow and wrist:	☐ Wrist extensor tendon	☐	☐	☐	☐
Shoulder:	☐ Subscapularis	☐	☐	☐	☐
	☐ Suprasinatus	☐	☐	☐	☐
	☐ Infraspinatus	☐	☐	☐	☐
	☐ Teres minor	☐	☐	☐	☐
Other: _____		☐	☐	☐	☐

Please tick which **ligament/s** you have injured? (next column on the tright)

Also indicate if your sprained or completely tore the ligamant

Ligament	L	R	Sprain	Complete Tear
☐ Shoulder ligaments	☐	☐	☐	☐
☐ Elbow ligaments	☐	☐	☐	☐
☐ Wrist ligaments	☐	☐	☐	☐
☐ Finger ligaments	☐	☐	☐	☐
☐ Knee (ACL)	☐	☐	☐	☐
☐ Knee (MCL)	☐	☐	☐	☐
☐ Knee (PCL)	☐	☐	☐	☐
☐ Knee (LCL)	☐	☐	☐	☐
☐ Ankle lateral ligaments	☐	☐	☐	☐
☐ Ankle medial ligaments	☐	☐	☐	☐
☐ Spinal ligaments	☐	☐	☐	☐
☐ Other: _____	☐	☐	☐	☐

Please tick if you have you ever suffered from any of the following **joint capsule** injuries?

- ☐ Acute shoulder dislocation
- ☐ Chronic shoulder instability
- ☐ Chronic ankle instability
- ☐ Other: _____

Do you suffer from any other **connective tissue or rheumatological diseases** or disorders? (If yes, please specify which one)

Yes ☐ No ☐ (refer to the list on the next page)
(If yes, specify: _____)

List of some Connective Tissue and/or Rheumatic Diseases and Disorders

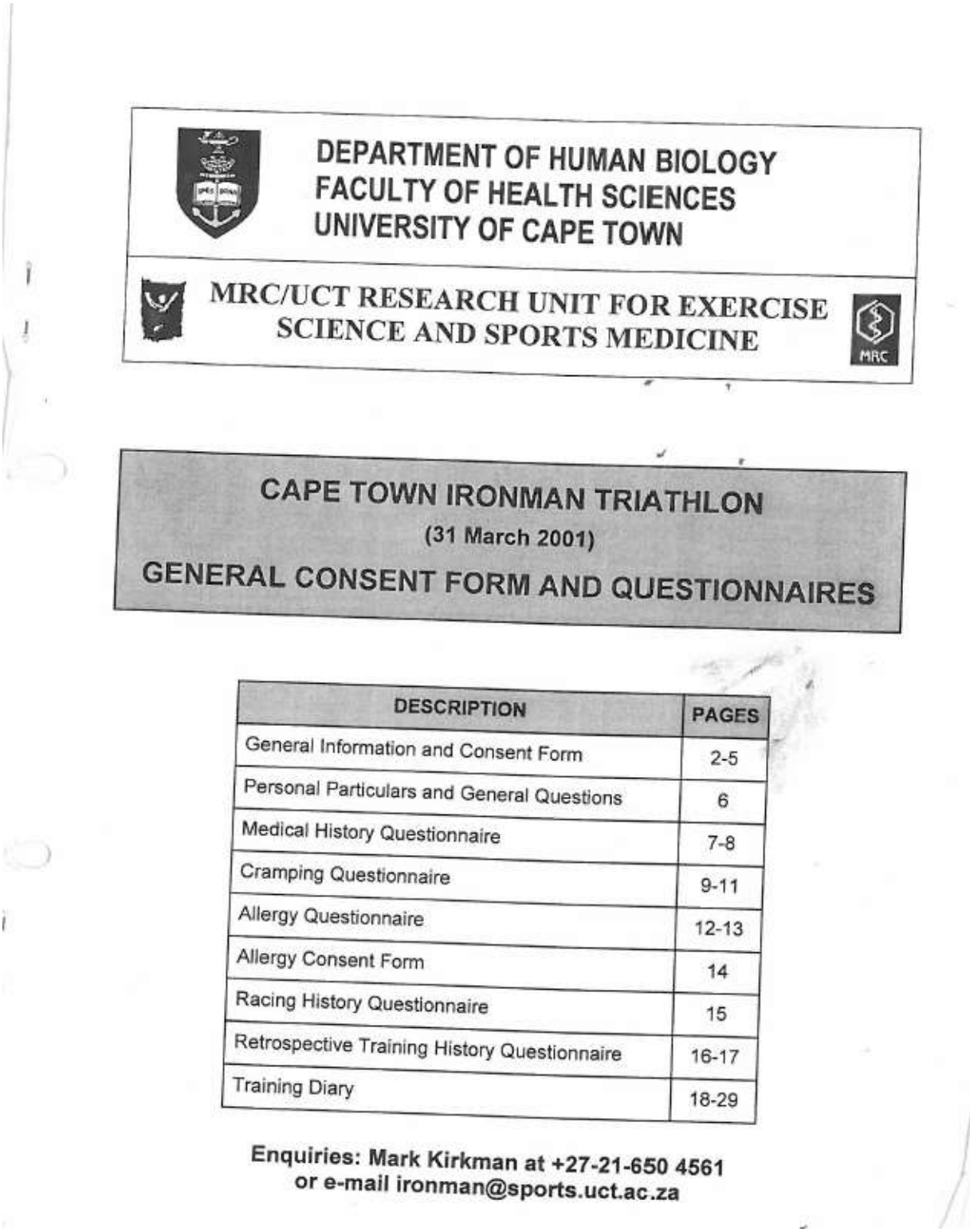
Ankylosing Spondylitis	Lipid Storage Diseases	Pseudogout
Aspartylglycosaminuria (AGU)	Marfan Syndrome	Reactive Arthritis
Behcet's Syndrome	Menkes Kinky Hair Syndrome	Reiter's Syndrome
Crohn's Disease	Mucopolysaccharidoses	Relapsing Polychrondritis
Discoid Lupus Erythematosus	Myopathies and Dystrophies	Scleroderma
Ehlers-Danlos syndrome (EDS)	Ochronosis (Homocystinuria)	Sjogren's Syndrome
Eosinophilic Fascitis	Osteogenesis imperfecta (OI)	Systemic Lupus Erythematosus (SLE)
Giant Cell (Temporal) Arthritis	Polyartentis Nodosa	Systemic Sclerosis
Gout	Polymyalgia Rheumatica	Wegener's Granulomatosis
Hypersentive Vasulatis	Polymyositis & Dermatomyositis	

CURRENT INJURY

Current Injury 1			
What was the approximate date when you first became aware of the injury?	Month	Year	
Please indicate which side of your body is injured (if applicable)	☐ Right	☐ Left	
Please indicate which anatomical area is currently injured	☐ Head ☐ Neck ☐ Face ☐ Front chest ☐ Back chest ☐ Shoulder ☐ Upper arm	☐ Elbow ☐ Forearm ☐ Wrist ☐ Finger ☐ Lower back ☐ Hip ☐ Thigh	☐ Hamstring ☐ Quadriceps ☐ Knee ☐ Shin ☐ Achilles ☐ Ankle ☐ Foot
	Other (Specify: _____)		
Please indicate the type of structure that was injured	☐ Muscle ☐ Tendon ☐ Bone	☐ Ligament ☐ Joint	
	Other (Specify: _____)		
Please indicate in which sport (discipline) the injury occurred	☐ Running ☐ Swimming	☐ Cycling	
	Other (Specify: _____)		
Please indicate the severity of the injury (tick one box please)	☐ I only experience symptoms after exercise - Grade 1 ☐ I experience symptoms during exercise, but it does not interfere with exercise - Grade 2 ☐ I experience symptoms during exercise that may interfere with my training/ competition - Grade 3 ☐ I am so painful that I may not be able to train or compete - Grade 4		
Please indicate how your injury was treated to date (you can tick more than one)?	☐ Rest ☐ Stretches ☐ Physiotherapy ☐ Surgery ☐ Strengthening exercises ☐ Equipment change	☐ Tablets ☐ Cortisone injection ☐ Other injection ☐ Orthotics	
	Other (Specify: _____)		

Current Injury 2		
What was the approximate date when you first became aware of the injury?		Month Year
Please indicate which side of your body is injured (if applicable)		☐ Right ☐ Left
Please indicate which anatomical area is currently injured		☐ Head ☐ Neck ☐ Face ☐ Front chest ☐ Back chest ☐ Shoulder ☐ Upper arm ☐ Elbow ☐ Forearm ☐ Wrist ☐ Finger ☐ Lower back ☐ Hip ☐ Thigh ☐ Hamstring ☐ Quadriceps ☐ Knee ☐ Shin ☐ Achilles ☐ Ankle ☐ Foot Other (Specify: _____)
Please indicate the type of structure that was injured		☐ Muscle ☐ Tendon ☐ Bone ☐ Ligament ☐ Joint Other (Specify: _____)
Please indicate in which sport (discipline) the injury occurred		☐ Running ☐ Cycling ☐ Swimming Other (Specify: _____)
Please indicate the severity of the injury (tick one box please)		☐ I only experience symptoms after exercise - Grade 1 ☐ I experience symptoms during exercise, but it does not interfere with exercise - Grade 2 ☐ I experience symptoms during exercise that may interfere with my training/ competition - Grade 3 ☐ I am so painful that I may not be able to train or compete - Grade 4
Please indicate how your injury was treated to date (you can tick more than one)?		☐ Rest ☐ Stretches ☐ Physiotherapy ☐ Surgery ☐ Strengthening exercises ☐ Equipment change ☐ Tablets ☐ Cortisone injection ☐ Other injection ☐ Orthotics Other (Specify: _____)

Questionnaire used at the 2000 and 2001 South African Ironman Triathlon. Data collected was used in Chapters 5-7. Range of Motion, Endurance Performance and Exercise-associated Muscle Cramps.



DEPARTMENT OF HUMAN BIOLOGY
FACULTY OF HEALTH SCIENCES
UNIVERSITY OF CAPE TOWN

MRC/UCT RESEARCH UNIT FOR EXERCISE
SCIENCE AND SPORTS MEDICINE

MRC

CAPE TOWN IRONMAN TRIATHLON

(31 March 2001)

GENERAL CONSENT FORM AND QUESTIONNAIRES

DESCRIPTION	PAGES
General Information and Consent Form	2-5
Personal Particulars and General Questions	6
Medical History Questionnaire	7-8
Cramping Questionnaire	9-11
Allergy Questionnaire	12-13
Allergy Consent Form	14
Racing History Questionnaire	15
Retrospective Training History Questionnaire	16-17
Training Diary	18-29

Enquiries: Mark Kirkman at +27-21-650 4561
or e-mail ironman@sports.uct.ac.za

GENERAL INFORMATION

The MRC/UCT Research Unit for Exercise Science and Sports Medicine will be conducting a research trial at the Cape Town IRONMAN Triathlon to be held in Gordons Bay on 31 March 2001. The research trial will focus on the following three aspects:

1. Monitor fluid and sodium balance in all triathletes.
2. Identify both genetic and non-genetic factors that predict performance ability during this ultra-endurance event.
3. Identify factors that predispose ultra-endurance athletes to cramps.

All triathletes taking part in the Cape Town IRONMAN Triathlon (31st March 2001) will be requested to participate in this research study. Athletes will however, not be coerced to take part in this study. As a participant in this research study, you will receive an in-depth questionnaire (pages 6 to 25 of this book) to complete prior to the triathlon. You will be requested to provide information on your personal particulars, your lifestyle and sporting history, and your medical history.

Those triathletes wishing to-participate in the additional studies (i.e. salt intake, energy expenditure, weight monitoring and fluid ingestion) may volunteer or enquire by contacting Mark Kirkman at +27-21-650 4561 or ironman@sports.uct.ac.za.

Registration

At race registration (1 - 4 days before the race), you will be requested to report to the University of Cape Town (UCT) "on-site" research venue which will be located near the registration tables at the Villa Via Hotel in Gordons Bay. At the UCT research venue, you will hand in your questionnaires and a researcher will check the completeness of the questionnaires with you.

You will then be requested to donate a 10-ml (2 teaspoons) blood sample. Half the sample will be used for genetic analysis (to predict performance ability) and the remainder will be used to determine your blood sodium concentration. All athletes will also be requested to be weighed, have their height measured, and have their blood pressure measured in a supine position (lying flat).

Race-day

At registration, you will be issued with a microchip (as used in other races), that you will attach to your ankle as a form of identification. This chip will be used to record the time taken to complete each stage of the race as well as your finishing time for the complete triathlon. In addition, the chip will be used to identify you when you are weighed so that your weight can be recorded automatically. You will be requested to weigh yourself before and after the race and have the option of weighing yourself at the transition points, and during each of the laps during the marathon. Changes in your body weight can be used to monitor your fluid balance during the race and will not interfere with your performance. Your weight change will also be used as a guide to the type of treatment that will be best for you should you require medical care either during or after the race.

Immediate post-race testing

On completion or termination of the triathlon, you will be issued with a medical card. You will then be requested to enter the medical tent situated near the finish line for a post-race medical examination regardless of whether or not you feel that you require medical care. It is important that all triathletes should be screened medically at the end of the race to insure that any incipient conditions are identified early. In addition it is important that data are collected on athletes who do not actively seek medical care at the finish so that their data can be used as control data for comparison with those triathletes who do require medical care. In this way your participation will be of value to all triathletes around the world, as it will help us better to understand the factors causing some triathletes to require medical care after the completion of these events.

During the medical examination you will be requested to donate another blood sample (5ml, 1 teaspoon) to assess the changes in your blood sodium concentration. This measurement helps to provide information of your fluid status at the finish of the race. Your rectal temperature will also be measured.

Triathletes requiring medical care during or after the race

If you fail to complete the triathlon for medical reasons or if you require medical care at the finish of the race, either because you collapse or develop symptoms that require medical attention, you will receive the appropriate medical care provided by physicians trained in sports medicine and in the care of the specific conditions likely to be present in IRONMAN triathletes. Those athletes that present with cramps may be requested by the physician to

have a muscle biopsy for research purposes. However, this will only be done if you consent to do so, you will not be coerced.

There is considerable uncertainty of the best ways to treat the common forms of collapse found in IRONMAN athletes. By measuring your weight before and after the race we will know the extent to which dehydration or alternatively over-hydration, is contributing to your condition. This will insure that you do not receive inappropriate therapy including either too little or too much fluid during the recovery period.

In addition, should your body temperature be abnormally elevated (above 41 degrees centigrade), we will place you in a bath of ice cold water until your body temperature falls below 38 degrees Centigrade.

If you should develop a more severe degree of lowered blood sodium concentration (hyponatraemia) so that you require admission to hospital, your urine production during recovery will be collected and sufficient blood samples will be drawn to insure that your recovery is expedited.

Testing on the days following the race.

The day following the race we will re-measure body weights of all the triathletes who are present at the prize giving ceremony in Gordons Bay. Triathletes who remain in the Cape Town area after the race will be encouraged to report as frequently as possible for re-weighing at the Sports Science Institute of South Africa in Newlands, preferably on a daily basis until Wednesday 4th April.

This information is important for us to determine how rapidly the body refills its fluid stores after the race. This will help us to decide whether all the weight lost during the race is actually due to fluid loss, that is required to be replaced during exercise. This, in turn, will help us ascertain how much fluid athletes should actually receive during medical treatment at the end of the IRONMAN.

PERSONAL PARTICULARS

(TO BE COMPLETED BY ALL PARTICIPANTS)

Event	Cape Town Ironman Triathlon	Race Number	
Surname			
First name(s)			
Postal address		Code	
E-mail address		Phone	
Date of birth		Cell	
Height (cm)	☐☐☐,☐ cm	Gender	male ☐ female ☐
Weight (kg)	☐☐☐,☐ kg		
Ethnic group (Only Required and Used for Research Purposes)	Black/African ☐ White ☐ Indian ☐ Mixed Ancestry (Coloured) ☐ Asian ☐ Other ☐		
Ancestry: Tribal or national background (eg Xhosa, Dutch, Irish)	Father	Unknown ☐	
	Mother	Unknown ☐	
Nationality			
Country of birth		Dominant Handedness	Left ☐ Right ☐
Smoker	Yes (Current) ☐ Yes (Ex smoker) ☐ No, never ☐		
	If yes, number of years _____	If stopped, when _____	
	If yes, number per day _____		
Do you know your blood group?	Yes ☐ A ☐ B ☐ AB ☐ O ☐		
	No ☐ Pos ☐ Neg ☐		
Have you ever had achilles tendon problems?	Yes ☐ No ☐		

BMI

Google M.

MEDICAL HISTORY QUESTIONNAIRE

(TO BE COMPLETED BY ALL PATICIPANTS)

Do you suffer from any medical conditions? e.g. Asthma, Diabetes, Heart Disease	Yes ☐ No ☐
	If Yes, Please Specify. Heart Disease ☐ Kidney Disease ☐ Lung Disease ☐ Diabetes ☐ Asthma ☐ Other ☐ Please specify:
Are you on any medications (Over the Counter or Prescription)?	Yes ☐ No ☐
	If Yes, Please specify:
Have you ever collapsed dunng a race?	Yes ☐ No ☐
	If Yes, What was the race distance?
	If Yes, What were the weather conditions (Temp, humidity, etc)?
	If Yes, What was your state of training?
	If Yes, What were your symptoms?
	If Yes, What was the reason for the collapse?
	If Yes, How were you treated?

Have you ever collapsed after a race?	Yes ☐ No ☐
	If Yes, What was the race distance?
	If Yes, What were the weather conditions (Temp, humidity, etc)?
	If Yes, What was your state of training?
	If Yes, What were your symptoms?
	If Yes, What was the reason for the collapse?
	If Yes, How were you treated?

MUSCLE CRAMPING QUESTIONNAIRE

(TO BE COMPLETED BY ALL PATICIPANTS)

1.Have you ever experienced muscle cramping during or immediately after exercise? (Please tick the box if you have)

a) in your running career ☐
b) in the last 5 years ☐
c) in the last year ☐

2.Have you ever experienced any other form of muscle cramping? (you can tick more than one box)

a) Night cramps ☐
b) Cramps during pregnancy (if applicable) ☐
c) Side stitch during exercise ☐
d) Other cramps ☐
Please specify_____

3.Please tick the box that shows which type of exercise where you are most likely to cramp? (Please tick only one box)

a) Running ☐
b) Swimming ☐
c) Cycling ☐
d) Other ☐
Please specify_____

4.Please tick under which conditions you experience cramping

a) only in training ☐
b) only in racing ☐
c) in training and racing ☐

5.Please complete this question if you ever experienced cramping during training sessions:

Please indicate how many times you have experienced cramping in the **last 10 training sessions** (Insert the number 0 to 10 in the box)

a) during running ☐
b) during swimming ☐
c) during cycling ☐

At what point during the training sessions do you usually first experience cramping

a) From the beginning ☐
b) In the first quarter of the training session ☐
c) In the second quarter of the training session ☐
d) In the third quarter of the training session ☐
e) In the last quarter of the training session ☐
f) After the training session ☐

6. Please complete this question if you ever experienced cramping during races:

Please indicate how many times you have experienced cramping in the **last 10 races** (Insert the number 0 to 10 in the box)

a) during running ☐
b) during swimming ☐
c) during cycling ☐

At what point during the race do you usually first experience cramping

a) From the beginning ☐
b) In the first quarter of the race ☐
c) In the second quarter of the race ☐
d) In the third quarter of the race ☐
e) In the last quarter of the race ☐
f) After the race ☐

7. Please indicate in which muscle groups do you cramp? Please rank each muscle group as follows 0=never cramp in the muscle, 2=sometimes cramp in the muscle, 3=usually cramp in the muscle

a) Quadriceps ☐
b) Hamstrings ☐
c) Calves ☐
d) Inner thigh ☐
e) Foot ☐
f) Other (please specify) _____ ☐

8. Please indicate how does the cramp last for (in minutes)? ☐ minutes

9. Please indicate the severity of your cramping using the following classification (please tick one box)

a) Mild: 5 minutes and you are able to continue exercising ☐
b) Moderate: 10 minutes and you are able to continue exercising ☐
c) Severe: 15> minutes and you have to STOP exercising ☐

10. Please indicate which of the following factors alleviate your cramping. Also, please rate the success of each factor in **alleviating your cramping** during exercise as follows (1=not successful, 2=moderately successful, 3=very successful)

Success rating in alleviating cramping

a) Immediate stretching ☐
b) Drinking fluids ☐
c) Intravenous fluids ☐
d) Resting the muscle ☐
e) Massage ☐
f) Ice massage ☐
g) Contracting the cramping muscle ☐
h) Contracting the opposite muscle (e.g. hamstring if the quad cramps) ☐
i) Contracting the muscle on the opposite side (left if right cramps) ☐
j) Medication (specify) _____ ☐
k) Other (specify) _____ ☐

11. Please indicate whether there is any one in your family that has ever experienced cramping during exercise or any other form of muscle cramping (e.g. night cramps)

	Exercise cramping	**Night cramping**	**Other cramping**
a) Father	☐	☐	☐
b) Mother	☐	☐	☐
c) Grandfather	☐	☐	☐
d) Grandmother	☐	☐	☐
e) Brother	☐	☐	☐
f) Sister	☐	☐	☐

12.Please indicate, by ticking the appropriate box, whether you have consulted a health professional regarding you cramping

a) Nutritionist/Dietician ☐
b) Doctor ☐
c) Physiotherapist ☐
d) Sports scientist ☐
e) Coach ☐
f) Podiatrist ☐
g) Chiropractor ☐
h) Other health professional ☐
Please specify_____

13.Please indicate, by ticking the appropriate box, what treatment you have had for your muscle cramping to date. Also, please rate the success of each treatment that you had in **preventing cramping** during exercise as follows (1=not successful, 2=moderately successful, 3=very successful)

	Success rating in preventing cramping
a) Regular stretching	☐
b) Increased training	☐
c) Decreased training	☐
d) Magnesium supplementation	☐
e) Sodium supplementation	☐
f) Increased fluid intake during exercise	☐
g) Decreased fluid intake during exercise	☐
h) Adequate race preparation	☐
i) Muscle strengthening program	☐
j) Massage therapy	☐
k) Use of supplements (specify type)_____	☐
l) Medication (specify type) _____	☐
m) Other (please specify)_____	☐

14. Which of the following factors are associated with your cramping during exercise (Please insert a number into each box that indicates the relative importance that you place on each of the listed factors being associated with your cramping: **1=not related at all, 2=possibly related, 3=probably related, 4=definitely related**)

	Rating as a factor associated with cramping (1-4)
a) previous muscle injury	☐
b) overhydration	☐
c) dehydration	☐
d) "flu like" symptoms before the race/training	☐
e) muscle "twitchiness" just before cramping	☐
f) muscle fatigue just before cramping	☐
g) not stretching the muscles regularly	☐
h) overstretching the muscles	☐
i) increasing your training load	☐
j) not using nutritional supplements	☐
k) genetic factors (family members)	☐
l) lack of electrolytes (minerals)	☐
m) lack of "fitness"	☐
n) hot weather	☐
o) Other (specify)_____	☐

ALLERGY QUESTIONAIRE

(TO BE COMPLETED BY ALL PATICIPANTS)

Do you have a family history of allergies in.	Parents	Yes ☐	No ☐
	Siblings	Yes ☐	No ☐
Did you have any infantile allergies such as atopic dermatitis (eczema) or food allergies?	Yes ☐	No ☐	
Did you have hay fever as a young child?	Yes ☐	No ☐	
Do you have hay fever currently?	Yes ☐	No ☐	
Have you had hay fever in the past?	Yes ☐	No ☐	
Are your symptoms seasonal?	Yes ☐	No ☐	
In which months of the year do you have nasal symptoms?	January ☐	February ☐	
	March ☐	April ☐	
	May ☐	June ☐	
	July ☐	August ☐	
	September ☐	October ☐	
	November ☐	December ☐	
What are your symptoms?	Sneezing ☐	Itchy palate ☐	
	Runny nose ☐	Itchy eyes ☐	
	Streaming eyes ☐	Blocking of nose ☐	
	Headache ☐	Fatigue ☐	
	Poor sleep ☐		
Do you have nasal allergy problems in winter which are not associated with "colds" or "flu"?	Yes ☐	No ☐	

Do you take medication for hay fever?	Yes ☐ No ☐
	If yes, do you take it: Daily ☐ Occasionally ☐ Only in the summer/spring ☐
If yes, specify which tablets or nasal sprays you take	
Have you had allergy tests before?	Yes ☐ No ☐
If you have confirmed allergies please indicate to which allergen	House dust mite ☐ Dogs ☐ Cats ☐ Grass pollen ☐ Tree pollen ☐ Moulds ☐ Foods ☐ Cockroach ☐
Others (specify)	
Do you have asthma?	Yes ☐ No ☐
What triggers your symptoms?	
Does exercise trigger your symptoms?	Yes ☐ No ☐
Are you on intermittent treatment?	Yes ☐ No ☐
Specify what treatment you take for your asthma	
Are you prepared to have a set of allergy skin tests and a blood test to determine whether you are at risk of developing allergies?	Yes ☐ No ☐

If you are prepared to have a set of allergy skin tests and a blood test please read and sign the consent form on the next page (page 14).

ALLERGY CONSENT FORM

(TO BE COMPLETED BY PATICIPANTS IN THE ALLERGY STUDY)

I _____ hereby consent to participate in the evaluation of allergies in Ironman Triathletes, conducted by the Research Unit for Exercise Science and Sports Medicine (Department of Human Biology, University of Cape Town) during registration at the Villa Via Hotel in Gordons Bay. I understand that I will be requested to give a blood sample (5ml, 1 teaspoon) and have skin tests performed on my arms.

Name of triathlete _____

Signature of triathlete _____

Date _____

Name of investigator _____

Signature of investigator _____

Date _____

RACING HISTORY

(TO BE COMPLETED BY ALL PARTICIPANTS)

NOTE: If you are unsure of what training or races you have completed or participated in, please do not guess but rather indicate that you are unsure or can't remember.

Type of triathlon	Sprint	Olympic	½ Ironman	Ironman
Which triathlons have you participated in?	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐
Personal best time (hrs:min:sec)				
Year of first event				
How many events have you participated in?				
Running events	**5 km**	**10 km**	**21.1 km**	**42.2 km**
Which running events have you participated in?	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐
Personal best time (hrs:min:sec)				
Year of first event				
How many events have you participated in?				

Please complete the following table to document the number and distance of the races that you have participated in over the past 5 years. Please distinguish between the type of race eg. triathlon or cycle tour etc. If you took part in triathlons, biathlons or duathlons, document the total distance of each discipline (S = swim, R = run, C = cycle).

Time	Triathlons		Biathlons		Duathlons		Swimming		Cycling		Running	
	number	Total distance	Number	Total distance	number	Total distance	number	Total distance	number	Total distance	number	Total distance
0-1 years		S: C: R:		S: R:		R: C:						
1-2 years		S: C: R:		S: R:		R: C:						
2-3 years		S: C: R:		S: R:		R: C:						
3-4 years		S: C: R:		S: R:		R: C:						
4-5 years		S: C: R:		S: R:		R: C:						

RETROSPECTIVE TRAINING HISTORY QUESTIONNAIRE
(TO BE COMPLETED BY ALL PARTICIPANTS)

The following questionnaire will provide us with information on your training over the past 5 years.

NOTE: *If you are unsure of what training or races you have completed or participated in, please do not guess but rather indicate that you are unsure or can't remember.*

Please complete the following tables to document your swimming, cycling and running training volume over the last 5 years.
Note: estimate the average hours and the average distance per week that you have trained in the 2nd and 3rd column.

SWIMMING **Age you started swimming (years):** ☐

Where applicable

Interval or high-intensity training included, yes or no

Injuries and how many days the injury prevented you from swimming

	Months/ year	Average hrs/wk	Average distance/wk	Interval training		Injuries		Comment
Time			(km)	Yes or no		Yes or no, Time no training		
0-3 months	—			yes	no	yes	no	
						days:		
3-6 months	—			yes	no	yes	no	
						days:		
6-12 months	—			yes	no	yes	no	
						days:		
1-2 years				yes	no	yes	no	
						days:		
2-3 years				yes	no	yes	no	
						days:		
3-4 years				yes	no	yes	no	
						days:		
4-5 years				yes	no	yes	no	
						days:		

CYCLING

Age you started cycling (years): ☐

Where applicable

Interval or high-intensity training included, yes or no

Injuries and how many days the injury prevented you from cycling

	Months/ year	Average hrs/wk	Average distance/wk	Interval training		Injuries		Comment
Time			(km)	Yes or no		Yes or no, Time no training		
0-3 months	—			yes	no	yes	no	
						days:		
3-6 months	—			yes	no	yes	no	
						days:		
6-12 months	—			yes	no	yes	no	
						days:		
1-2 years				yes	no	yes	no	
						days:		
2-3 years				yes	no	yes	no	
						days:		
3-4 years				yes	no	yes	no	
						days:		
4-5 years				yes	no	yes	no	
						days:		

RUNNING

Age you started running (years): ☐

Where applicable

Interval or high-intensity training included, yes or no

Injuries and how many days the injury prevented you from running

	Months/ year	Average hrs/wk	Average distance/wk	Interval training		Injuries		Comment
Time			(km)	Yes or no		Yes or no, Time no training		
0-3 months	—			yes	no	yes	no	
						days:		
3-6 months	—			yes	no	yes	no	
						days:		
6-12 months	—			yes	no	yes	no	
						days:		
1-2 years				yes	no	yes	no	
						days:		
2-3 years				yes	no	yes	no	
						days:		
3-4 years				yes	no	yes	no	
						days:		
4-5 years				yes	no	yes	no	
						days:		

17

TRAINING DIARY

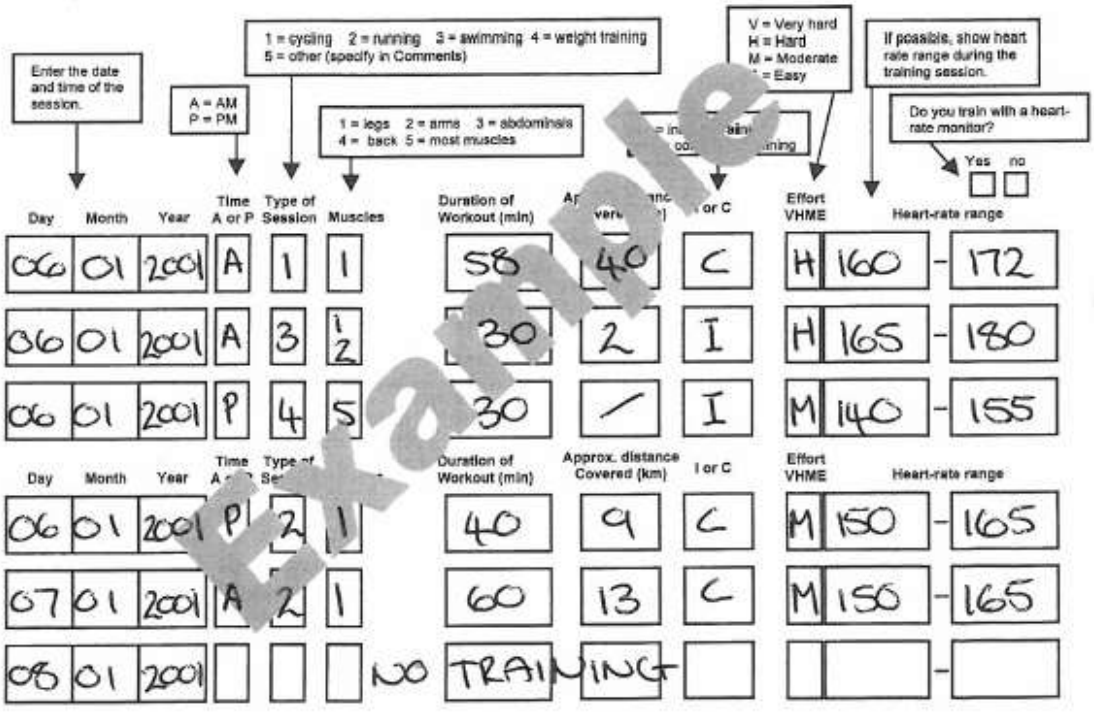
(TO BE COMPLETED BY ALL PARTICIPANTS)

Please complete the training diary for a ***month*** prior to the Ironman Triathlon.

- You must document every day that you train (or don't train)
- Use one block (row) for each training session.
- If you train more than once a day or train more than one discipline at a time, document each of these in a separate block (row)
- For each session (there can be more than one per day), please record the:
  - date,
  - time (morning or afternoon/evening),
  - type of session (running, swimming, cycling, weight training, other)
  - which muscle groups were mainly used,
  - time and distance (if applicable) of the session,
  - if the session involved interval training,
  - the effort used (very hard, hard, moderate, easy)
  - your heart-rate range during the session.
- Please specify if you use a heart-rate monitor or you monitor your heart-rate manually.
- Refer to the example on the first page as a reference.

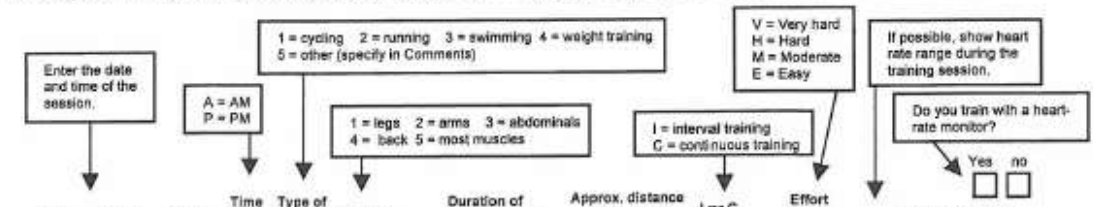
TRAINING DIARY

Use one row of boxes for each type of workout (you can have a number of workouts on one day). Put any comments in the box at the bottom of the page. Link the comments to the sessions.



TRAINING DIARY

Use one row of boxes for each type of workout (you can have a number of workouts on one day). Put any comments in the box at the bottom of the page. Link the comments to the sessions.



Day	Month	Year	Time A or P	Type of Session	Muscles	Duration of Workout (min)	Approx. distance Covered (km)	I or C	Effort VHME	Heart-rate range
										–
										–
										–

Day	Month	Year	Time A or P	Type of Session	Muscles	Duration of Workout (min)	Approx. distance Covered (km)	I or C	Effort VHME	Heart-rate range
										–
										–
										–

Day	Month	Year	Time A or P	Type of Session	Muscles	Duration of Workout (min)	Approx. distance Covered (km)	I or C	Effort VHME	Heart-rate range
										–
										–
										–

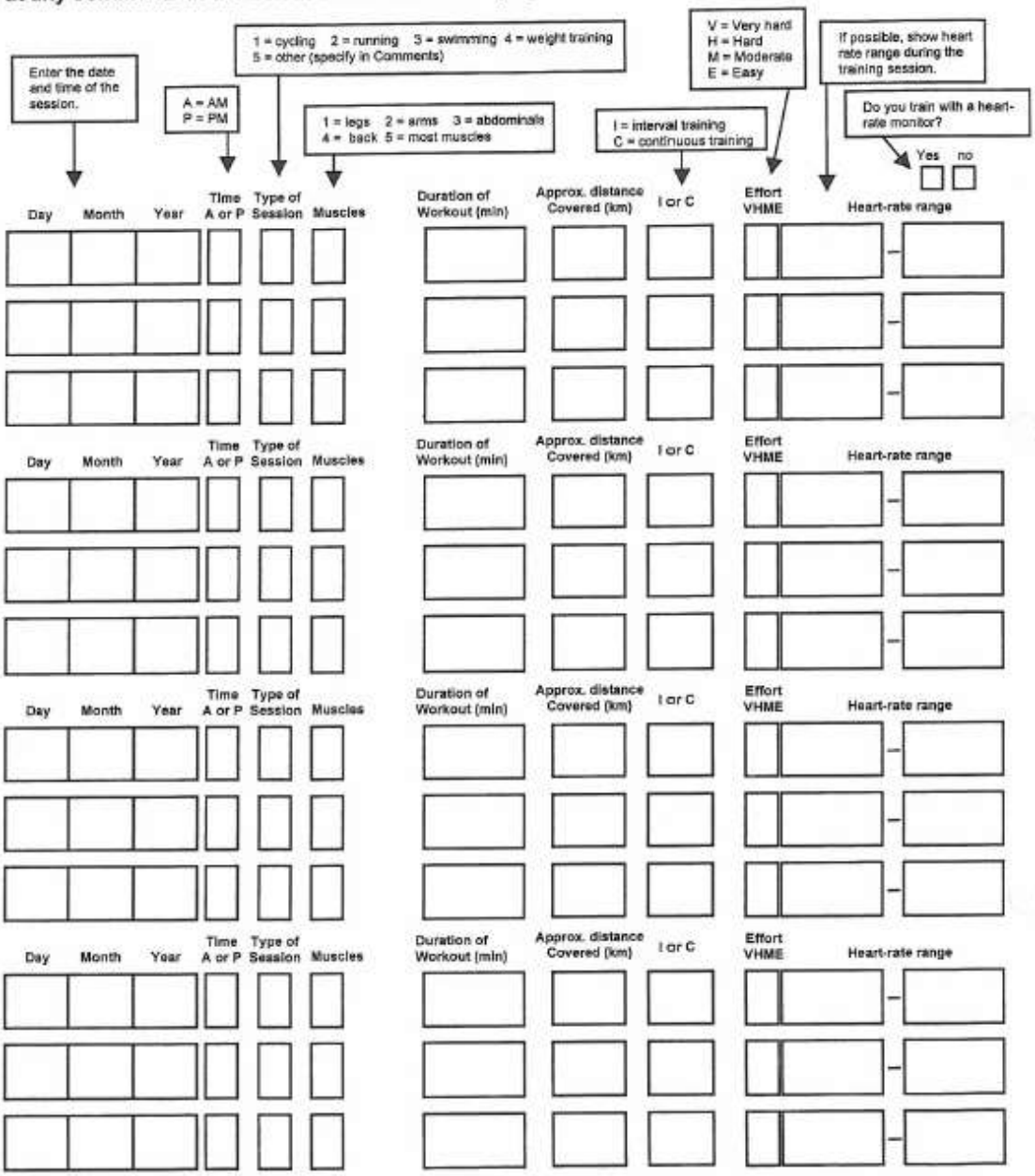
Day	Month	Year	Time A or P	Type of Session	Muscles	Duration of Workout (min)	Approx. distance Covered (km)	I or C	Effort VHME	Heart-rate range
										–
										–
										–

Comments

19

TRAINING DIARY

Use one row of boxes for each type of workout (you can have a number of workouts on one day). Put any comments in the box at the bottom of the page. Link the comments to the sessions.



Questionnaire used at the 2006 and 2007 South African Ironman Triathlon. Data collected was used in Chapters 5-7. Range of Motion, Endurance Performance and Exercise-associated Muscle Cramps.

Department of Human Biology
UCT/MRC RESEARCH UNIT FOR EXERCISE SCIENCE & SPORTS MEDICINE
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: + 27 21 650 4561
Fax: + 27 21 686 7530

2007 IRONMAN – MEDICAL AND TRAINING QUESTIONNAIRES

These questionnaires have been constructed by the Medical Research team, in conjunction with the Medical Director of the Ironman 2007. The information obtained from these questionnaires is essential for the planning of medical care during events such as the Ironman. We acknowledge that the questionnaires are long, but we are asking about 30 minutes of your valuable time to complete them. The completion of the questionnaires is voluntary; all the information will be kept confidential and will only be used for research and medical care planning purposes. We suggest that you consider downloading and completing this before the event and handing in the completed questionnaire, at the research area during race registration.

Prof Martin Schwellnus (Chairman, Research Team)
Dr Peter Schwartz (Medical Director, Ironman 2007)

Instructions

Please answer each question by filling in the details in the allocated space or checking one or more of the option boxes.

Please bring the completed forms together with the signed consent form to the research table at race registration.

Please complete sections A, B, C, D, E and F

Please complete only the relevant questions in the following section

Section A: Personal details			
2007 Ironman Race Number			
Surname			
First Name			
Postal Address		Postal/ Zip Code	
E-mail address		Phone (day time)	code number
Alternate E-mail address			
Date of birth	yyyy-mm-dd	Cell (Mobile)	
Height	cm	Gender	Male ☐ Female ☐
Weight	kg	Age (on race day)	yrs
Ethnic group (Only Required and Used for Research Purposes)	Black/African ☐ White ☐ Indian ☐ Mixed Ancestry (Coloured) ☐ Asian ☐ Other ☐		
Ancestry: Tribal or national background (eg Xhosa, Dutch, Zulu, German, Italian)	Father: Unknown ☐		
	Mother: Unknown ☐		
Country of Birth			
Dominant Hand	Left ☐ Right ☐ Both ☐	Dominant Leg	Left ☐ Right ☐ Both ☐
Occupation			
What **percentage** of your **working** day is spent in the following activities?	Sitting: ____ % Standing: ____ % Walking (Lower body activity) ____ % Manual Labour (upper and body activity) ____ %		
Did you participate in the research project conducted at the 2006 Ironman in Port Elizabeth			Yes ☐ No ☐

Section B. Racing and training history

Type of triathlon	Standard (1.6, 40, 10)	Ironman	
Which triathlons have you **ever** participated in?	Yes ☐ No ☐	Yes ☐ No ☐	
Year of first event			
How many triathlon events have you **ever** participated in?			
How many triathlon races have you completed over the **past 2 years**?			
Personal best time **ever**	_____ hrs:min	_____ hrs:min	
What was your time for your last triathlon race during the **past 12 months**?	_____ hrs:min	_____ hrs:min	
Type of running event	**10 km**	**21.1 km**	**42.2 km**
Which road running races have you **ever** participated in?	Yes ☐ No ☐	Yes ☐ No ☐	Yes ☐ No ☐
Year of first event			
How many events have you **ever** participated in?			
Personal best time **ever**	_____ min	_____ min	_____ min
What is your best time, in a running race, in the **last 15 weeks**?	_____ min	_____ min	_____ min
Type of event	**Two Oceans Marathon**	**Comrades Marathon**	
Which races have you **ever** participated in?	Yes ☐ No ☐	Yes ☐ No ☐	
Year of first event			
How many events have you **ever** participated in?			
Personal best time	_____ hrs:min	_____ hrs:min	
What is your personal best cycling time in a race between **80 to 120 km** in the **last 15 weeks**?	Time: _____ min Distance: _____ km		
South African Ironman Triathlon racing history Did you enter any of the South African Ironman Triathlons?			
2000 (Gordon's Bay)	Yes ☐ No ☐ Race No _____		
2001 (Gordon's Bay)	Yes ☐ No ☐ Race No _____		
2005 (Port Elizabeth)	Yes ☐ No ☐ Race No _____		
2006 (Port Elizabeth)	Yes ☐ No ☐ Race No _____		
What is your predicted time for the entire 2007 Ironman event and each of the three splits?	Entire event: _____ min Swim: _____ min Cycle: _____ min Run: _____ min		

Please answer the following questions, with your answers reflecting your average in the **most recent 15 weeks i.e. beginning December 2006 to 18th March, 2007.**	
Do you train with a heart rate monitor?	Yes ☐ No ☐
Do you race with a heart rate monitor?	Yes ☐ No ☐
Do you record, download and store your heart rate information?	Yes ☐ No ☐
Would you be willing to make your heart rate data available to the research team?	Yes ☐ No ☐
How many days a week did you train during the **last 15 weeks**?	days/week
What distances did you train in an average week during the **last 15 weeks**?	Swim: _____ km/week Cycle: _____ km/week Run: _____ km/week
How many hours a week did you **train** in an average week during the **last 15 weeks**?	Swim: _____ hrs/week Cycle: _____ hrs/week Run: _____ hrs/week
How many hours a week did you **work** in an average week during the **last 15 weeks**?	_____ hrs/week
What **distances** did you train in the week before the race?	Swim: _____ km Cycle: _____ km Run: _____ km
How many **hours** did you train in the **week before** the race?	Swim: _____ hours Cycle: _____ hours Run: _____ hours
How many fast/ hard sessions did you do per week in the **last 8 weeks**?	Swim: _____ Cycle: _____ Run: _____
Describe briefly the session, including distance, time and recovery interval (if applicable) e.g. 10 x 400m in 75 sec with 60 sec jog recovery between each	
What percentage of your weekly training distance was done at race speed or faster (for each discipline)?	Swim: _____ % Cycle: _____ % Run: _____ %
How many hours did you train 3 days before the race	Swim: _____ hours Cycle: _____ hours Run: _____ hours
How many hours did you train 2 days before the race	Swim: _____ hours Cycle: _____ hours Run: _____ hours
How many hours did you train the day before the race	Swim: _____ hours Cycle: _____ hours Run: _____ hours
How did your training commitment affect your social life?	☐ Not at all ☐ A fair amount ☐ A lot

Flexibility training history

Do you perform flexibility training (regular stretching exercises)?	Yes ☐ No ☐
If YES, please complete the rest of the flexibility training history section below:- If NO, continue completing the questionnaire from the top of page 5 (Equipment use history).	
On average, how many days a week do you perform a stretching session?	days/week
On average, how times a day do you perform a stretching session?	times/day
Please tick which muscle groups do you include in your stretching session?	☐ Hamstrings ☐ Quadriceps ☐ Calf (gastrocnemius) ☐ Calf (soleus) ☐ Groin (inner thigh) ☐ Upper body limbs ☐ Other: _____
Please tick when you stretch? (before, during and/or after exercising. You can tick more than one box)	☐ Before Exercise ☐ During Exercise ☐ After Exercise
When you stretch an individual muscle group, on average, **how long do you hold the stretch** for?	seconds
When you stretch an individual muscle group, on average, **how many times do you stretch the muscle for**?	☐ Once ☐ Twice ☐ 3 times ☐ 4 times ☐ 5 times ☐ 6 or more times

Equipment use history

Please indicate which type of **bicycle** you use?	☐ Kuota ☐ Aegis ☐ Felt ☐ Cervelo ☐ Elite ☐ Giant	☐ Kestrel ☐ Litespeed ☐ Quintana Roo ☐ Argon 18 ☐ Specialized ☐ Other: _____	☐ Trek ☐ Softride ☐ Javelin ☐ Scott ☐ Guru
Please indicate which type of **handle bars** you use?	☐ Bontrager ☐ Profile Design ☐ Deda ☐ Pedalsoft ☐ Other: _____	☐ HED ☐ Vision Tech ☐ Easton ☐ Kestrel	☐ Zipp ☐ Oval Concepts ☐ Syntace

Please indicate which type of **saddle** (Brand - model) you use?	☐ Selle San Marco- Azoto TriathGel ☐ Profile Design- Tri Stryke (with a groove) ☐ Selle San Marco- Rever Profil ☐ Fizik- Arione Tri ☐ Terry ☐ Koobi ☐ Other: _____
Please indicate which brand of **helmet** you use?	☐ Trek ☐ Bell ☐ Giro ☐ MET ☐ Other: _____
Please indicate which type of **cycling shorts** you use?	☐ Thin lycra (no padding) ☐ Padded cycling shorts ☐ Triathlon shorts with some padding ☐ Swimming costume ☐ Other: _____
Do you normally wear **underwear** together with cycling shorts?	☐ Yes ☐ No
Please indicate which type of **cycling shoes** you use?	☐ Olympic ☐ Nike ☐ Diadora ☐ Shimano ☐ Carnac ☐ Sidi ☐ Other: _____
Please indicate which type of **kit** you use?	☐ Anatomic ☐ Nike ☐ Velo ☐ Howzit ☐ Adidas ☐ Orca ☐ De Soto ☐ Louis Garneau ☐ Quintana Roo ☐ Zoot ☐ Other: _____
Please indicate which **brand of running shoe** you use?	☐ Adidas ☐ Asics ☐ Brooks ☐ New Balance ☐ Nike ☐ Mizuno ☐ Puma ☐ Reebok ☐ Saucony ☐ Other: _____
Please indicate which **type of running shoe** you use?	☐ Soft neutral shoe ☐ Mild anti-pronation shoe ☐ Motion control shoe ☐ Light racing shoe ☐ Unknown or not sure ☐ Other: _____

Section C. History of medication and supplement use

What medication, if any, are you currently using? (please list)	Name of medication	Years taken

Do you use protective skin sunscreen during training session or when competing?	Yes ☐ No ☐	☐ Every session	☐ Most sessions
		☐ Some sessions	☐ Very occasionally

Are you currently taking dietary supplements/vitamins? Yes ☐ No ☐

If **yes** to the above question, please list names of dietary, sports or vitamin supplements.	Name of supplement	Years taken
	☐ Multi-vitamins	____
	☐ Anti-oxidants	____
	☐ Immune boosters	____
	☐ Protein powders/supplements, Protein bars. BCAAs	____
	☐ Creatine	____
	☐ Caffeine	____
	☐ Fat cutters	____
	☐ Carbohydrate drinks/powders/gels	____
	☐ Other: _____	____

Have you ever used oral corticosteroids (cortisone tablets)? (If **yes**, how long ago?)	Yes ☐ No ☐	☐ 3 months	☐ 6 months
		☐ 12 months	☐ 24 or more months
Have you ever been given an injection with corticosteroids? (If **yes**, how long ago?)	Yes ☐ No ☐	☐ 3 months	☐ 6 months
		☐ 12 months	☐ 24 or more months
Have you ever been given an injection of corticosteroids in or around the **Achilles** tendon? (If **yes**, how many times?)	Yes ☐ No ☐	☐ Once	☐ Twice
		☐ 3 times	☐ >3 times
Have you ever used fluoroquinoline antibiotics? (refer to the following list)	Yes ☐ No ☐	☐ 3 months	☐ 6 months
		☐ 12 months	☐ 24 or more months

List of some fluoroquinoline antibiotics:

ADCO-CIPRIN	CIPROBAY	SANDOZ CIPROFLOXACIN
AVELON	CIPROGEN	TAFLOC
BACTIDRON	CPL ALLIANCE CIPROFLOXACIN	TARIVID
CIFLOC	DYNAFLOC	TAVANIC
CIFRAN	FACTIVE	TEQUIN
CIPLA-CIPROFLOXACIN	FLOXIN	UNIQUIN
CIPLOXX	MAXAQUIN	UTIN-400
CIPRO-HEXAL	NOROXIN	ZANOCIN
	ORPIC	

Lifestyle and habits history			
Please indicate your smoking status	Current smoker ☐	Ex smoker ☐	Never smoked ☐
If you answered yes, (past or current smoker) please complete the section on the right	Number of years of smoking:	If stopped, how many years ago:	
	What is (was) the average number of cigarettes per day:		
On average, how much alcohol do you drink per week (tots, glasses) of spirits, wine or beer?	_____ glasses beer/cider per week _____ glasses wine per week _____ tots of spirits per week		

Fluid Intake					
How do you best describe your fluid intake during an Ironman triathlon race?	(a) I drink to thirst ☐ (b) I drink as much as tolerable ☐ (c) I drink according to a predetermined fluid intake schedule ☐ (d) I drink to prevent any weight loss during exercise ☐ (e) I combine (a) with (c) ☐ (f) I combine (b) with (c) ☐ (g) Other: _____ ☐				
What percentage of your fluid intake will consist of these beverages?	Water:	☐ 0-25%	☐ 26-50%	☐ 51-75%	☐ 76-100%
	Sports drink:	☐ 0-25%	☐ 26-50%	☐ 51-75%	☐ 76-100%
	Coke:	☐ 0-25%	☐ 26-51%	☐ 51-75%	☐ 76-100%
	Other:	☐ 0-25%	☐ 26-50%	☐ 51-75%	☐ 76-100%
	Specify other: _____				
What will be your estimated <u>total</u> fluid intake be **(if at all)** during the **<u>swim</u>**?	ml				
What will be your estimated <u>total</u> fluid intake be during the **<u>cycle</u>**?	ml				
What will be your estimated <u>total</u> fluid intake be during the **<u>run</u>**?	ml				
Rank the following sources of information on their importance in formulating your drinking strategy. (1 being most influential and the lowest number being least influential)	_____ Fellow triathletes _____ Coach / trainer _____ Magazines / books _____ Website (please specify: _____) _____ Drinking guidelines from sports associations _____ Adverts _____ Self-experimentation _____ Other: _____				

Section D. Psychological and Behavioural

Connor-Davidson Resilience Scale (CD-RISC)

Please indicate how much you agree with the following statements as they apply to you over the last **month**. If a particular situation has not occurred recently, answer according to how you think you would have felt.

	not true at all	rarely true	sometimes true	often true	true nearly all the time
1. I am able to adapt when changes occur.					
2. I have at least one close and secure relationship which helps me when I am stressed.					
3. When there are no clear solutions to my problems, sometimes fate or God can help.					
4. I can deal with whatever comes my way.					
5. Past successes give me confidence in dealing with new challenges and difficulties.					
6. I try to see the humorous side of things when I am faced with problems.					
7. Having to cope with stress can make me stronger.					
8. I tend to bounce back after illness, injury, or other hardships.					
9. Good or bad, I believe that most things happen for a reason.					
10. I give my best effort, no matter what the outcome may be.					
11. I believe I can achieve my goals, even if there are obstacles.					
12. Even when things look hopeless, I don't give up.					
13. During times of stress/crisis, I know where to turn for help.					
14. Under pressure, I stay focused and think clearly.					
15. I prefer to take the lead in solving problems, rather than letting others make all the decisions.					
16. I am not easily discouraged by failure.					
17. I think of myself as a strong person when dealing with life's challenges and difficulties.					
18. I can make unpopular or difficult decisions that affect other people, if it is necessary.					
19. I am able to handle unpleasant or painful feelings like sadness, fear and anger.					
20. In dealing with life's problems, sometimes you have to act on a hunch, without knowing why					
21. I have a strong sense of purpose in life.					
22. I feel in control of my life.					
23. I like challenges.					
24. I work to attain my goals, no matter what roadblocks I encounter along the way.					
25. I take pride in my achievements.					

TPQ / TCI (96 shared items)		
1. I usually am confident that everything will go ell, even in situations that worry most people.	True☐	False ☐
2. I often try new things just for fun or thrills, even if most people think it is a waste of time.	True☐	False ☐
3. I like to discuss my experiences and feelings openly with friends instead of keeping them to myself.	True☐	False ☐
4. When nothing new is happening, I usually start looking for something that is thrilling or exciting.	True☐	False ☐
5. Usually I am more worried about that most people that something might go wrong in the future.	True☐	False ☐
6. I don't mind discussing my personal problems with people whom I have known briefly or slightly.	True☐	False ☐
7. I would like to have warm and close friends with me most of the time.	True☐	False ☐
8. I nearly always stay relaxed and carefree even when nearly everyone else is fearful.	True☐	False ☐
9. I usually demand very good practical reasons before I am willing to change my old ways of doing things.	True☐	False ☐
10. I often have to stop what I am doing because I start worrying that something might go wrong.	True☐	False ☐
11. I hate to change the way I do things, even if many people tell me there is a new and better way to do it.	True☐	False ☐
12. My friends find it hard to know my feelings because I seldom tell them about my private thoughts.	True☐	False ☐
13. I like it when people can do exactly what they want without strict rules and regulations.	True☐	False ☐
14. I often stop what I am doing because I get worried, even when my friends tell me everything will go well.	True☐	False ☐
15. It wouldn't bother me to be alone all the time.	True☐	False ☐
16. I like to be very organized and set up rules for people whenever I can.	True☐	False ☐
17. I usually do things my own way, rather than giving in to the wishes of other people.	True☐	False ☐
18. I usually feel tense and worried when I have to do something new and unfamiliar.	True☐	False ☐
19. I often feel tense and worried in familiar situations, even when others feel there is little to worry about.	True☐	False ☐
20. Other people often think that I am too independent because I won't so what they want.	True☐	False ☐
21. Even when most people feel it is not important, I often insist on things being done in a strict and orderly way.	True☐	False ☐
22. I often do things based on how I feel at the moment, without thinking about how they are done in the past.	True☐	False ☐
23. I often feel tense and worried in unfamiliar situations, even when others feel there is no danger at all.	True☐	False ☐
24. I often brake rules and regulations when I think I can get away with it.	True☐	False ☐
25. I don't care very much whether other people like me or the way I do things.	True☐	False ☐
26. I usually stay calm and secure in situations that most people would find physically dangerous.	True☐	False ☐
27. I feel it is more important to be sympathetic and understanding of other people than to be practical and tough- minded.	True☐	False ☐
28. I lose my temper more quickly than most people.	True☐	False ☐
29. I am usually confident that I can easily do things that most people would consider dangerous (such as driving as automobile fast on a wet or icy road).	True☐	False ☐

30. I often react so strongly to unexpected news that I say or do things that I regret.	True☐	False ☐
31. People find it easy to come to me for help, sympathy, and warm understanding.	True☐	False ☐
32. I am much more reserved and controlled than most people.	True☐	False ☐
33. When I have to meet a group of strangers, I am more shy than most people.	True☐	False ☐
34. I am strongly moved by sentimental appeals (like when asked to helped crippled people).	True☐	False ☐
35. I almost never get so excited that I lose control of myself.	True☐	False ☐
36. I have a reputation as someone who is practical and does not act on emotion.	True☐	False ☐
37. I often avoid meeting strangers because I lack confidence with people I do not know.	True☐	False ☐
38. I usually stay away from social situations where I would have to meet strangers, even if I am assured that they will be friendly.	True☐	False ☐
39. I usually push myself harder than most people do because I want to do as well as I possibly can.	True☐	False ☐
40. I often push myself to the point of exhaustion or try to do more than I really can.	True☐	False ☐
41. I would probably stay relaxed and outgoing when meeting a group of strangers, even if I were told they were unfriendly.	True☐	False ☐
42. It is difficult for me to keep the same interests for a long time because my attention often shifts to something else.	True☐	False ☐
43. I think I would stay confident and relaxed when meeting strangers, even if I were told they are angry with me.	True☐	False ☐
44. I could probably accomplish more than I do, but I don't see the point of pushing myself harder than is necessary to get by.	True☐	False ☐
45. I like to think about things for a long time before I make a decision.	True☐	False ☐
46. Most of the time I would prefer to do something a little risky (like riding in an automobile over steep hills and sharp turns), rather than having to stay quiet and inactive for a few hours.	True☐	False ☐
47. I often follow my instincts, hunches, or intuition without thinking through all the details.	True☐	False ☐
48. I try to do as little work as possible, even when other people expect more of me.	True☐	False ☐
49. I often have to change my decisions because I had a wrong hunch or mistaken first impression.	True☐	False ☐
50. Most of the time I would prefer to so something risky (like hang-gliding or parachute jumping(, rather than having to stay quiet and inactive for a few hours.	True☐	False ☐
51. I am satisfied with my accomplishments and have little desire to do better.	True☐	False ☐
52. I see no point in continuing to work on something unless there is a good chance of success.	True☐	False ☐
53. I have less energy and get tired more quickly than most people.	True☐	False ☐
54. I usually think about all the facts in detail before I make a decision.	True☐	False ☐
55. I nearly always think about all the facts in detail before I make a decision, even when other people demand a quick decision.	True☐	False ☐
56. I often need naps or extra rest periods because I get tired so easily.	True☐	False ☐
57. I don't go out of my way to please other people.	True☐	False ☐
58. I am more energetic and tire less quickly than most people.	True☐	False ☐
59. I am usually able to get other people to believe me, even when I know that what I am saying is exaggerated or untrue.	True☐	False ☐
60. I can usually do a good job of stretching the truth to tell a funnier story or to play a joke on someone.	True☐	False ☐
61. I usually can stay "on the go" all day without having to push myself.	True☐	False ☐

62. I am usually more upset than most people by the loss of a close friend.	True ☐	False ☐
63. I have trouble telling a lie, even when it is meant to spare someone else's feelings.	True ☐	False ☐
64. I am better at saving money than most people.	True ☐	False ☐
65. Even after there are problems in a friendship, I nearly always try to keep it going anyway.	True ☐	False ☐
66. I recover more slowly than most people from minor illnesses or stress.	True ☐	False ☐
67. I need much extra rest, support, or reassurance to recover from minor illnesses or stress.	True ☐	False ☐
68. I often spend money until I run out of cash or get into debt from using too much credit.	True ☐	False ☐
69. Because I so often spend too much money on impulse, it is hard for me to save money, even for special plans like a vacation.	True ☐	False ☐
70. It is extremely difficult for me to adjust to changes in my usual way of doing things because I get so tense, tired or worried.	True ☐	False ☐
71. If I am feeling upset, I usually feel better around friends than when left alone.	True ☐	False ☐
72. I usually feel much more confident and energetic than most people, even after minor illnesses or stress.	True ☐	False ☐
73. Some people think I am too stingy or tight with my money.	True ☐	False ☐
74. I often keep trying the same thing over and over again, even when I have not had success in a long time.	True ☐	False ☐
75. It is hard for me to enjoy spending money on myself, even when I have saved plenty of money.	True ☐	False ☐
76. I recover more quickly than most people from minor illnesses or stress.	True ☐	False ☐
77. I hate to make decisions based only on my first impressions.	True ☐	False ☐
78. I think I will have very good luck in the future.	True ☐	False ☐
79. I am most often moved deeply by fine speech or poetry.	True ☐	False ☐
80. If I am embarassed or humiliated, I get over it very quickly.	True ☐	False ☐
81. I like old "tried and true" ways of doing things according to their priority of importance to me because of lack of time.	True ☐	False ☐
82. I like to keep my problems to myself.	True ☐	False ☐
83. I enjoy saving money more than spending it on entertainment or thrills.	True ☐	False ☐
84. Even when I am with friends, I prefer not to "open up" very much.	True ☐	False ☐
85. I feel very confident and sure of myself in almost all social situations.	True ☐	False ☐
86. I usually like to stay cool and detached from other people.	True ☐	False ☐
87. I never worry about terrible things that might happen in the future.	True ☐	False ☐
88. I am more hard-working than most people.	True ☐	False ☐
89. In conversations I am much better as a listener than as a talker.	True ☐	False ☐
90. I like to please other people as much as I can.	True ☐	False ☐
91. Regardless of any temporary problem that I have to overcome, I always think it will turn out well.	True ☐	False ☐
92. I like to stay at home better than to travel and explore new places.	True ☐	False ☐
93. I am usually so determined that I continue to work long after other people have given up.	True ☐	False ☐
94. I usually have good luck in whatever I try to do.	True ☐	False ☐
95. I like to pay close attention to details in everything I do.	True ☐	False ☐
96. It is easy for me to organize my thoughts while talking to someone.	True ☐	False ☐

2007 Ironman Race No: _____

K10

Instructions: The following questions ask about how you have been feeling during the **past four weeks**. For each question, please circle the number that best describes how often you have had this feeling. Your answers will be kept confidential.

In the past four weeks:	None of the time	A little of the time	Sometime of the time	Most of the time	All of the time
1. About how often did you feel tired of for no good reason?	1	2	3	4	5
2. About how often did you feel nervous?	1	2	3	4	5
3.About how often did you feel so nervous that nothing could calm you down?	1	2	3	4	5
4. About how often did you feel hopeless?	1	2	3	4	5
5. About how often did you feel restless or fidgety?	1	2	3	4	5
6. About how often did you feel restless you could not sit still?	1	2	3	4	5
7. About how often did you feel depressed?	1	2	3	4	5
8. About how often did you feel that everything is an effort?	1	2	3	4	5
9. About how often did you feel so sad that nothing could cheer you up?	1	2	3	4	5
10. About how often did you feel worthless?	1	2	3	4	5

2007 Ironman Race No: _____

Section E. Family medical history

Have any of your blood (biological) relatives ever had the following?

Please tick yes or no. If yes, please tick the relationship of that person to you (You may tick more than one of the relationship blocks).

Description		If Yes, please indicate the relationship
Exercise associated muscle cramps	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Night muscle cramps	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Chronic Achilles tendon injury	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Achilles tendon rupture	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Any ligament injury	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Asthma	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Allergies (in general)	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Heart Disease	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Diabetes	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Depression, Anxiety attacks, Personality disorder	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Gastro-intestinal (GIT) disease	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother

2007 Ironman Race No: _____

Section F. Personal general medical history

In this section, you are asked to read through 14 questions about your personal general medical history. If you answer "yes" to any of questions 1 to 12, please complete the additional questions at the end of the section (section G on page 18).

1. In the **6 weeks before this race** (from 1st February) did you suffer from any **symptoms of flu** (fever, sore throat, blocked or runny nose, cough, wheeze, muscle aches and pains)?	Yes ☐ No ☐
2. Have you **ever** in triathlon career suffered from **muscle cramping** (painful, spontaneous, sustained spasm of a muscle) during or immediately (within 6 hours) after exercise (in training or competition)?	Yes ☐ No ☐
3. Have you **ever** in your triathlon career suffered from **a tendon or ligament injury** (pain, swelling, stiffness) in any tendon (including Achilles tendon, knee tendons, and shoulder tendons) or ligaments (partial or complete tear)?	Yes ☐ No ☐
4. Have you **ever** in your triathlon career **used medicines to treat injuries** in the week **before or during a race** – including anti-inflammatory drugs, cortisone (pills, or injection), or pain killers?	Yes ☐ No ☐
5. Have you **ever** in your triathlon career suffered **gastrointestinal** symptoms **during exercise** including heartburn, nausea, vomiting, abdominal pain, urge to defecate (pass a stool), diarrhoea, or blood in the stools?	Yes ☐ No ☐
6. Have you **ever** in your triathlon career suffered from symptoms of the **nervous system** including exercise induced headaches, nerve tingling or loss of sensation?	Yes ☐ No ☐
7. Have you **ever** in your triathlon or cycling career (in particular with **cycling**) suffered from **injury to the genital area** including genital numbness after cycling, genital pain after cycling, genital swelling or altered sexual function after cycling?	Yes ☐ No ☐
8. Have you **ever** in your triathlon career suffered from **symptoms of allergies** including nose allergies (hay fever), allergic sinusitis, allergic asthma, skin allergies, a past history of allergies to medication, plant material or animal material?	Yes ☐ No ☐
9. Do you **currently suffer from asthma** including exercise induced asthma, or symptoms of asthma such as shortness of breath, wheezing, or chronic coughing?	Yes ☐ No ☐
10. Have you ever **collapsed** (fell down **not because of an accident**, needing medical attention) during, at the finish or after a race or training session?	Yes ☐ No ☐
11. Do you **currently** suffer from any **symptoms of injury** in the muscles, tendons, bones, ligaments or joints?	Yes ☐ No ☐
12. Do you **currently**, or did you **in the last year**, suffer from any symptoms of **exercise related skin disease**?	Sunburn: Yes ☐ No ☐ Skin cancer: Yes ☐ No ☐ Other skin damage resulting sun exposure: Yes ☐ No ☐

13. Please tick in which anatomical area you ever had surgery performed.	☐ Gastric (stomach) ☐ Oesophageal (swallowing pipe) ☐ Small bowel ☐ Large bowel (colon) ☐ Rectum ☐ Gallbladder ☐ Pancreas ☐ Liver ☐ Abdomen (general) ☐ Wrist ☐ Head ☐ Finger ☐ Neck ☐ Lower back ☐ Face ☐ Hip ☐ Front chest ☐ Thigh ☐ Back chest ☐ Knee ☐ Shoulder ☐ Lower leg ☐ Upper arm ☐ Achilles ☐ Elbow ☐ Ankle ☐ Forearm ☐ Foot ☐ Other (Specify: _____)
14. Management of pain during the last 3 months	
14a. Did you alter or stop your training schedule due to pain in any part of your body?	Yes ☐ No ☐
If yes: For how long	_____ days
Did you adapt your training schedule for a while when your injury/illness was healed?	Yes ☐ No ☐
14b. How do you feel when you experience pain? (you can tick more than one option)	☐ It does not bother me much ☐ Angry ☐ Frustrated ☐ Depressed ☐ Resentful ☐ Overwhelmed
14c. When you experience pain, do you? (you can tick more than one option)	☐ Adjust your training schedule ☐ Stop training ☐ Slowly get "back on track" of your training schedule ☐ Train harder to make up for the missed training sessions ☐ Ignore the pain and continue to train ☐ Feel scared to do anything that could aggravate the pain ☐ Think that the pain means that you have a severe injury ☐ Tell everybody about it
15. **Female athletes only:** Please complete the following questions (14a. to 14g.) related to your menstrual cycle and other gynaecological history	
15a. At what age did you start your periods (menstruating)?	(years)
15b. In the last 12 months, how many menstrual cycles did you have?	
15c. Have you ever had irregular menstrual periods in the past? (excluding pregnancy)?	Yes ☐ No ☐

2007 Ironman Race No: _____

15d. Have you had a hysterectomy/ovarectomy?	Yes ☐ No ☐
15e. How many times have you been pregnant?	(times)
15f. What form of contraception are you currently using?	☐ None ☐ Oral contraceptive pill ☐ Injection ☐ Intra-uterine device ☐ Sterilization (tubes tied) ☐ Other: _____
15g. If yes to question 15f. above, for oral contraceptive pill, for what reason was the pill prescribed?	☐ Not applicable ☐ Dermatological ☐ Contraception ☐ Regulate period ☐ Other: _____

THANK YOU FOR COMPLETING THIS QUESTIONNAIRE

If you have answered **YES** to any of the first 11 questions of the Personal General Medical History questionnaire (section F) please complete the relevant additional questions that follow in section G.

Please bring the completed forms together with the signed consent form to the pre-race facility or the research table at race registration.

2007 Ironman Race No: _____

Section G. Additional detailed medical history

(Please complete all the sections to which you answered "Yes" in the Personal general medical history)

1. Flu symptoms in the last 6 weeks

If you answered **YES** to **question 1** in section F, please complete the following two questions related to flu symptoms in the last 6 weeks.

(1a) Please tick which of these flu symptoms you suffered from **in the last 6 weeks**.	☐ Fever ☐ Cough ☐ Joint pains ☐ Blocked nose ☐ Wheezing ☐ Sore Throat ☐ Runny nose ☐ Muscle aches ☐ Any other flu symptoms (Specify: _____)
(1b) Please tick which of these flu symptoms you suffered from **in the last 7 days**.	☐ Fever ☐ Cough ☐ Joint pains ☐ Blocked nose ☐ Wheezing ☐ Sore Throat ☐ Runny nose ☐ Muscle aches ☐ Any other flu symptoms (Specify: _____)

2. Muscle cramping

If you answered **YES** to **question 2** in section F, please complete the following questions (2a. to 2m.) related to your cramping.

(2a) For how many years have you suffered from cramping?	(years)
(2b) Did you suffer from cramping during or after exercise in the **last 12 months**?	Yes ☐ No ☐
(2c) With what **type of exercise** is your cramping associated (You can tick more than one form of exercise)?	☐ Swimming ☐ Cycling ☐ Running
(2d) In the **last 10 races or training sessions**, how many times have you experienced cramping?	Races: _____/10 Training sessions: _____/10
(2e) What treatment/s have you had that **successfully relieved** an acute cramp? (can tick more than one)	☐ Stretching ☐ Resting ☐ Drinking fluid ☐ Ice application ☐ Massage ☐ Magnesium ☐ Salt (tablets or solution) ☐ Other (Specify: _____)
(2f) At **what point in the race or training run** do you usually first experience cramping?	☐ First quarter ☐ Second quarter ☐ Third quarter ☐ Fourth quarter ☐ After the race ☐ No pattern
(2g) In which **muscles** do you usually cramp (please list the muscle by the one which cramps most frequently (as 1) and the others after that (2-4)?	☐ Calves ☐ Hamstrings ☐ Quadriceps (thigh) ☐ Foot muscles ☐ Other (Specify: _____)
(2h) Have you **ever** suffered from cramping in your **whole body** (arms and legs)?	Yes ☐ No ☐

2007 Ironman Race No: _____

(2i) Have you **ever** been **admitted to hospital** following cramping?	Yes ☐ No ☐
(2j) Have you **ever** been **confused or in a coma** during or after a cramping episode?	Yes ☐ No ☐
(2k) Have you ever had **"dark urine"** in the 3 days following a cramping episode?	Yes ☐ No ☐
(2l) If you cramp, **how long** does the cramp usually last for (min)?	(minutes)
(2m) If you cramp, how **severe** is the cramp usually? (please tick).	☐ Mild: < 5 minutes and you are able to continue exercising ☐ Moderate: 5-15 minutes and you are able to continue exercising ☐ Severe: >15 minutes or if you have to STOP exercising

3. Past Tendon and Ligament Injury History

If you answered **YES** to **question 3** in section F, please complete the following questions (3a. to 3d.) related to your past history of tendon/ligament injury/ies.

(3a) Please tick which **tendon/s** you have injured? (next column on the tright)

Also indicate (tick) if your injured tendon was longsatnding pain (tendinopathy) or an acute tear/rupture

	Tendon	Longstanding Pain (Tendinopathy)	Acute Tear/ Rupture
Foot and ankle:	☐ Achilles tendon	☐	☐
	☐ Tibialis posterior	☐	☐
	☐ Plantar fascia	☐	☐
Knee:	☐ Patellar tendon	☐	☐
Elbow and wrist:	☐ Wrist extensor tendon	☐	☐
Shoulder:	☐ Rotator cuff	☐	☐
Other: _____		☐	☐

(3b) Please tick which **ligament/s** you have injured? (next column on the tright)

Also indicate if your sprained or completely tore the ligamant.

Ligament	Sprain	Complete Tear
☐ Shoulder ligaments	☐	☐
☐ Elbow ligaments	☐	☐
☐ Wrist ligaments	☐	☐
☐ Finger ligaments	☐	☐
☐ Knee (ACL)	☐	☐
☐ Knee (MCL)	☐	☐
☐ Knee (PCL)	☐	☐
☐ Knee (LCL)	☐	☐
☐ Ankle lateral ligaments	☐	☐
☐ Ankle medial ligaments	☐	☐
☐ Spinal ligaments	☐	☐
☐ Other: _____	☐	☐

(3c) Please tick if you have you ever suffered from any of the following **joint capsule** injuries?	☐ Acute shoulder dislocation ☐ Chronic shoulder instability ☐ Other: _____
(3d) Do you suffer from any other **connective tissue or rheumatological diseases** or disorders? (If yes, please specify which one)	Yes ☐ No ☐ (refer to the list on the next page) (If yes, specify: _____)

2007 Ironman Race No: _____

List of some Connective Tissue and/or Rheumatic Diseases and Disorders		
Ankylosing Spondylitis	Lipid Storage Diseases	Pseudogout
Aspartylglycosaminuria (AGU)	Marfan Syndrome	Reactive Arthritis
Behcet's Syndrome	Menkes Kinky Hair Syndrome	Reiter's Syndrome
Crohn's Disease	Mucopolysaccharidoses	Relapsing Polychrondritis
Discoid Lupus Erythematosus	Myopathies and Dystrophies	Scleroderma
Ehlers-Danlos syndrome (EDS)	Ochronosis (Homocystinuria)	Sjogren's Syndrome
Eosinophilic Fasciitis	Osteogenesis imperfecta (OI)	Systemic Lupus Erythematosus (SLE)
Giant Cell (Temporal) Arthritis	Polyarteritis Nodosa	Systemic Sclerosis
Gout	Polymyalgia Rheumatica	Wegener's Granulomatosis
Hypersentive Vasulatis	Polymyositis & Dermatomyositis	

4. Use of medicines to treat an injury before or during participation

If you answered **YES** to **question 4** in section F, please complete the following two questions related to medicine use for injuries before or during races.

(4a) Which of the following medicines have you used in the past to treat an injury **in the week just before** a race?	☐ Paracetamol (e.g. Panado, Tylenol) ☐ Non-steroidal anti-inflammatories (e.g. Voltaren, Cataflam) ☐ Cortisone (pills) ☐ Cortisone injection ☐ Codeine ☐ Anti-inflammatory gels/creams/patches ☐ Any other pain killers (Specify: _____)
(4b) Which of the following medicines have you used in the past to treat an injury **during a race**?	☐ Paracetamol (e.g. Panado, Tylenol) ☐ Non-steroidal anti-inflammatories (e.g. Voltaren, Cataflam) ☐ Cortisone (pills) ☐ Cortisone injection ☐ Codeine ☐ Anti-inflammatory gels/creams/patches ☐ Any other pain killers (Specify: _____)

2007 Ironman Race No: _____

5. Gastrointestinal symptoms during exercise

If you answered YES to question 5 in section F, please indicate which gastrointestinal symptoms you have ever suffered from during exercise and, how frequently (in the last 12 months and in the last 10 races), and in which type of exercise.

Symptom	Number of times you experienced the GIT symptom in the last 12 months (during exercise)	Number of times you experienced the GIT symptom in the last 10 races (during races)	Please indicate which type of exercise is mostly associated with the GIT symptom	Please indicate the "severity" of the GIT symptom during exercise
Nausea			☐ Swimming ☐ Cycling ☐ Running	☐ Does not affect training or racing ☐ Affects training/racing (slow down or reduce time) ☐ Prevents training/racing
Vomiting			☐ Swimming ☐ Cycling ☐ Running	☐ Does not affect training or racing ☐ Affects training/racing (slow down or reduce time) ☐ Prevents training/racing
Heartburn			☐ Swimming ☐ Cycling ☐ Running	☐ Does not affect training or racing ☐ Affects training/racing (slow down or reduce time) ☐ Prevents training/racing
Abdominal pain			☐ Swimming ☐ Cycling ☐ Running	☐ Does not affect training or racing ☐ Affects training/racing (slow down or reduce time) ☐ Prevents training/racing
Urge to pass a stool (defecate)			☐ Swimming ☐ Cycling ☐ Running	☐ Does not affect training or racing ☐ Affects training/racing (slow down or reduce time) ☐ Prevents training/racing
Diarrhoea			☐ Swimming ☐ Cycling ☐ Running	☐ Does not affect training or racing ☐ Affects training/racing (slow down or reduce time) ☐ Prevents training/racing
Passing blood in the stool			☐ Swimming ☐ Cycling ☐ Running	☐ Does not affect training or racing ☐ Affects training/racing (slow down or reduce time) ☐ Prevents training/racing
Please indicate if you previously suffered from or had any of the following (you may tick more than one)?				☐ History of heartburn ☐ Gastroscopy ☐ Ulcer (gastric, duodenal) ☐ Irritable bowel syndrome ☐ Allergy to milk products ☐ Other past history of GIT disease

6. Diseases of the nervous system

If you answered YES to question 6 in section F, please indicate which nervous disease symptoms you have ever suffered from during exercise and, how frequently (in the last 12 months and in the last 10 races), and in which type of exercise.

Symptom	Number of times in the last 12 months (during exercise)	Number of times in last 10 races (during races)	Tick type of exercise
Headaches			☐ Swimming, ☐ Cycling, ☐ Running
Nerve tingling in the hands			☐ Swimming, ☐ Cycling, ☐ Running
Loss of sensation in the hands			☐ Swimming, ☐ Cycling, ☐ Running

2007 Ironman Race No: _____

7. Genital tract injury during cycling

If you answered YES to question 7 in section F, please indicate which symptoms of genital tract injury have you suffered from during or after cycling, how frequently (in the last 10 sessions), how long symptoms last, and what factors prevent or relieve symptoms?

Symptom	Number of times in the last 10 cycling sessions	Please indicate when the symptoms occur	Please indicate if any of the following reduce or prevent the symptoms (can tick more than one)
Genital numbness		☐ Only during cycling ☐ During and up to 1 hour after cycling ☐ During and 1-24 hours after cycling ☐ During and > 24 hours after cycling	☐ Changing the saddle type ☐ Changing the saddle position ☐ Using padded cycling shorts ☐ Wearing no underwear ☐ Wearing additional underwear ☐ Other (Specify: _____)
Genital pain		☐ Only during cycling ☐ During and up to 1 hour after cycling ☐ During and 1-24 hours after cycling ☐ During and > 24 hours after cycling	☐ Changing the saddle type ☐ Changing the saddle position ☐ Using padded cycling shorts ☐ Wearing no underwear ☐ Wearing additional underwear ☐ Other (Specify: _____)
Genital bruising		☐ Only during cycling ☐ During and up to 1 hour after cycling ☐ During and 1-24 hours after cycling ☐ During and > 24 hours after cycling	☐ Changing the saddle type ☐ Changing the saddle position ☐ Using padded cycling shorts ☐ Wearing no underwear ☐ Wearing additional underwear ☐ Other (Specify: _____)
Altered sexual function following a cycling session		☐ Up to 1 hour after cycling ☐ 1-24 hours after cycling ☐ > 24 hours after cycling	☐ Changing the saddle type ☐ Changing the saddle position ☐ Using padded cycling shorts ☐ Wearing no underwear ☐ Wearing additional underwear ☐ Other (Specify: _____)

2007 Ironman Race No: _____

8. Allergy history

If you answered **YES** to **question 8** in section F, please complete the following questions (8a. to 8e.) related to your current and past history of allergies.

(8a) Please indicate how long (years) have you been suffering from allergies?	years

(8b) Please tick which type of allergy do you currently suffer from

Nose (hay fever)	Yes ☐ No ☐	Sinusitis	Yes ☐ No ☐	Asthma (allergic)	Yes ☐ No ☐
Skin allergies	Yes ☐ No ☐	Eye allergies	Yes ☐ No ☐	Allergy to plant material	Yes ☐ No ☐
Allergy to foods	Yes ☐ No ☐	Allergy to animals	Yes ☐ No ☐	Allergy to medication	Yes ☐ No ☐

(8c) Please tick which type of allergy do you currently take medication for

Nose (hay fever)	Yes ☐ No ☐	Sinusitis	Yes ☐ No ☐	Asthma (allergic)	Yes ☐ No ☐
Skin allergies	Yes ☐ No ☐	Eye allergies	Yes ☐ No ☐	Allergy to plant material	Yes ☐ No ☐
Allergy to foods	Yes ☐ No ☐	Allergy to animals	Yes ☐ No ☐	Allergy to medication	Yes ☐ No ☐

(8d) Please tick which type of medication do you currently take

Cortisone nose spray	Yes ☐ No ☐	Cortisone nose inhaler	Yes ☐ No ☐	Anti-histamine tablets	Yes ☐ No ☐
Cortisone cream	Yes ☐ No ☐	Anti-histamine cream	Yes ☐ No ☐	Other inhaler / tablets or cream	Yes ☐ No ☐

(8e) Please tick which symptoms of allergy do you currently suffer from

Sneezing	Yes ☐ No ☐	Itchy runny nose	Yes ☐ No ☐	Headache	Yes ☐ No ☐
Itchy palate	Yes ☐ No ☐	Streaming eyes	Yes ☐ No ☐	Fatigue	Yes ☐ No ☐
Itchy eyes	Yes ☐ No ☐	Blocked nose	Yes ☐ No ☐	Poor sleep	Yes ☐ No ☐
Post nasal drip	Yes ☐ No ☐	Coughing	Yes ☐ No ☐	Wheezing	Yes ☐ No ☐
In which months of the year do you currently have symptoms of allergies? (You tick more than one)	☐ Jan ☐ Feb ☐ March ☐ April ☐ May ☐ June ☐ July ☐ Aug ☐ Sept ☐ Oct ☐ Nov ☐ Dec				

(8f) Please tick which type of allergy did you suffer from in the past (NOT currently)

Nose (hay fever)	Yes ☐ No ☐	Sinusitis	Yes ☐ No ☐	Asthma (allergic)	Yes ☐ No ☐
Skin allergies	Yes ☐ No ☐	Eye allergies	Yes ☐ No ☐	Allergy to plant material	Yes ☐ No ☐
Allergy to foods	Yes ☐ No ☐	Allergy to animals	Yes ☐ No ☐	Allergy to medication	Yes ☐ No ☐

9. Asthma history

If you answered **YES** to **question 9** in section F, please complete the following questions (9a. to 9k.) related to your current history of asthma

(9a) Do you currently suffer from asthma?	Yes ☐ No ☐
(9b) How many years have you suffered from asthma?	(years)
(9c) How was your asthma diagnosed?	☐ A doctor taking a history and performing an examination ☐ Lung function test (blow test) but no exercise ☐ Lung function test (blow test) before and after exercise ☐ Metacholine challenge test ☐ Eucapnic hyperventilation test (rebreathing test) ☐ Other test (Specify: _____)
(9d) Which <u>type of asthma</u> do you currently suffer from?	☐ Asthma that occurs at any time but <u>not during exercise</u> ☐ Asthma that occurs at any time including during exercise ☐ Asthma that <u>only</u> occurs <u>during exercise</u>
(9e) Please indicate **how frequently do you currently experience the symptoms** of asthma (shortness of breath, wheezing, coughing or coughing after exercise)?	**Daytime symptoms (per week)** ☐ < 2 / week ☐ 2-4 / week ☐ >4 / week ☐ All the time **Night time symptoms (per month)** ☐ < 1 / month ☐ 2-3 / month ☐ ≥4 / month ☐ All the time **Exercise related symptoms (per 10 exercise sessions)** ☐ <1 per 10 sessions ☐ 2-3 per 10 sessions ☐ ≥4 per 10 sessions
(9f) Please indicate if you had symptoms of asthma that were severe enough to necessitate **hospital admission in the last 12 months**	☐ No hospital admission for asthma in the last 12 months ☐ 1-2 hospital admissions for asthma in the last 12 months ☐ 3-4 hospital admissions for asthma in the last 12 months ☐ >4 hospital admissions for asthma in the last 12 months
(9g) Which <u>symptoms of asthma</u> do you currently suffer from?	☐ Wheezing ☐ Dry cough ☐ Shortness of breath ☐ Tight chest ☐ Chest pain ☐ Other (Specify: _____)

2007 Ironman Race No: _____

(9h) What **medication do you currently use** for your asthma? (you may tick more than one option)	☐ Cortisone inhaler (e.g. Beclate, Becloforte, Becodisks, Becotide, Budeflam, Flixotide, Inflammide, Pulmicort, Qvar, etc) ☐ Salbutamol (bronchodilator) inhaler (e.g. Ventolin, Venteze, Vomax, Airomir, Asthavent etc.) ☐ Salmeterol (bronchodilator) inhaler (Serevent) ☐ Fenoterol (bronchodilator) inhaler (Berotec) ☐ Terbutaline (bronchodilator) inhaler (Bricanyl) ☐ Formoterol (bronchodilator) inhaler (e.g. Foradil, Foratec, Oxis) ☐ Ipratropium (bronchodilator) inhaler (Atrovent) ☐ Tiotropium (bronchodilator) inhaler (Spiriva) ☐ Combined cortisone and bronchodilator inhaler (e.g. Atrovent, Berodual, Combivent, Duolin, Duovent, Seretide, Symbicord) ☐ Cortisone tablets ☐ Bronchodilator tablets ☐ Leukotriene receptor antagonist tablets (e.g. Acccolate, Singulair) ☐ Other inhaler ☐ Other medication (Specify: _____)
(9i) **When do you use your medication** for your asthma?	☐ Daily (irrespective of exercise) ☐ Only before exercise ☐ Other (Specify: _____)
(9j) **How long before an exercise session** do you use your medication for asthma?	min
(9k) Have you obtained **TUE (therapeutic use exemption forms)** for your asthma medication?	Yes ☐ No ☐

2007 Ironman Race No: _____

10. History of previous collapse If you answered **YES** to **question 10** in section F, please complete the following questions (10a. to 10d.) related to your current history of asthma	
(10a) Have you ever collapsed during training or racing?	☐ Training ☐ Racing ☐ Training and racing
(10b) How many times have you collapsed in training session or races during the last **five years**?	_____ training session _____ races
(10c) How many times have you collapsed in training session or races during the last **12 months** (1 year)?	
(10d) When you collapse, does it mostly occur before of after the finish line / completion of the training session?	☐ Before the finish ☐ After the finish
(10e) What is the cause of you collapse?	☐ Dehydration ☐ Heat illness ☐ Hyponatremia ☐ Low blood pressure ☐ Low blood sugar ☐ Other condition (Specify: _____)

2007 Ironman Race No: _____

11. History of any current injury that you suffer from

If you answered **YES** to **question 11** in section F, please complete the following questions (11a. to 11g.) related to each of your current injury/ies (Space is provided for two injuries)

Injury 1	
(11a) What was the approximate date when you first became aware of the injury?	Month Year
(11b) Please indicate which side of your body is injured (if applicable)	☐ Right ☐ Left
(11c) Please indicate which anatomical area is currently injured	☐ Head ☐ Neck ☐ Face ☐ Front chest ☐ Back chest ☐ Shoulder ☐ Upper arm ☐ Elbow ☐ Forearm ☐ Wrist ☐ Finger ☐ Lower back ☐ Hip ☐ Thigh ☐ Hamstring ☐ Quadriceps ☐ Knee ☐ Shin ☐ Achilles ☐ Ankle ☐ Foot Other (Specify: _____)
(11d) Please indicate the type of structure that was injured	☐ Muscle ☐ Tendon ☐ Bone ☐ Ligament ☐ Joint Other (Specify: _____)
(11e) Please indicate in which sport (discipline) the injury occurred	☐ Running ☐ Cycling ☐ Swimming Other (Specify: _____)
(11f) Please indicate the severity of the injury (tick one box please)	☐ I only experience symptoms after exercise - Grade 1 ☐ I experience symptoms during exercise, but it does not interfere with exercise - Grade 2 ☐ I experience symptoms during exercise that may interfere with my training/ competition - Grade 3 ☐ I am so painful that I may not be able to train or compete - Grade 4
(11g) Please indicate how your injury was treated to date (you can tick more than one)?	☐ Rest ☐ Stretches ☐ Physiotherapy ☐ Surgery ☐ Strengthening exercises ☐ Equipment change ☐ Tablets ☐ Cortisone injection ☐ Other injection ☐ Orthotics Other (Specify: _____)

2007 Ironman Race No: _____

Injury 2			
(11a) What was the approximate date when you first became aware of the injury?	Month	Year	
(11b) Please indicate which side of your body is injured (if applicable)	☐ Right	☐ Left	
(11c) Please indicate which anatomical area is currently injured	☐ Head ☐ Neck ☐ Face ☐ Front chest ☐ Back chest ☐ Shoulder ☐ Upper arm	☐ Elbow ☐ Forearm ☐ Wrist ☐ Finger ☐ Lower back ☐ Hip ☐ Thigh	☐ Hamstring ☐ Quadriceps ☐ Knee ☐ Shin ☐ Achilles ☐ Ankle ☐ Foot
	Other (Specify: _____)		
(11d) Please indicate the type of structure that was injured	☐ Muscle ☐ Tendon ☐ Bone	☐ Ligament ☐ Joint	
	Other (Specify: _____)		
(11e) Please indicate in which sport (discipline) the injury occurred	☐ Running ☐ Swimming	☐ Cycling	
	Other (Specify: _____)		
(11f) Please indicate the severity of the injury (tick one box please)	☐ I only experience symptoms after exercise - Grade 1 ☐ I experience symptoms during exercise, but it does not interfere with exercise - Grade 2 ☐ I experience symptoms during exercise that may interfere with my training/ competition - Grade 3 ☐ I am so painful that I may not be able to train or compete - Grade 4		
(11g) Please indicate how your injury was treated to date (you can tick more than one)?	☐ Rest ☐ Stretches ☐ Physiotherapy ☐ Surgery ☐ Strengthening exercises ☐ Equipment change	☐ Tablets ☐ Cortisone injection ☐ Other injection ☐ Orthotics	
	Other (Specify: _____)		

Questionnaire used for data collection for Chapter 9. Functional *COL6A1* rs35796750

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: + 27-21-650-4561 Fax: + 27-21-686-7530

DETERMINING THE FUNCTIONAL ROLE OF VARIANTS WITHIN THE EXTRACELLULAR MATRIX GENES ON MUSCULOSKELETAL SOFT TISSUE INJURIES, USING PRIMARY HUMAN FIBROBLAST CELL LINES.

Instructions

Please answer each question by filling in the details in the allocated space or checking one or more of the option boxes.

Please complete all twelve sections A to L

Subject Number: _____

Version 1
(March 2010)

Subject No: _____

Section A: Personal details			
Surname			
First Name			
Postal Address		Postal/ Zip Code	
E-mail address			
Alternate E-mail address		Phone (day time)	code number
Date of birth	yyyy-mm-dd	Cell (Mobile)	
Height	cm	Sex	Male ☐ Female ☐
Weight	kg	Age	yrs
Ethnic group (Only Required and Used for Research Purposes)	Black/African ☐ White ☐ Indian ☐ Mixed Ancestry (Coloured) ☐ Asian ☐ Other ☐		
Ancestry: Tribal or national background (eg Xhosa, Dutch, Zulu, German, Italian)	Father: Unknown ☐ Mother: Unknown ☐		
Country of Birth			
Dominant Hand	Left ☐ Right ☐ Both ☐	Dominant Leg	Left ☐ Right ☐ Both ☐
Current Occupation			
What **percentage** of your **working** day is spent in the following activities?	Sitting: _____ % Standing: _____ % Walking (Lower body activity) _____ % Manual Labour (upper and body activity) _____ %		
Occupation prior to muscle injury?			
Prior to injury, did your occupation involve lower or upper limb activity?	Yes ☐ No ☐		
If yes please indicate which arms/legs.	Right arm ☐ Left arm ☐ Both arms ☐ Right leg ☐ Left leg ☐ Both legs ☐		

Subject No: _____

Section B. Sporting Details Please record your sporting activities in order of importance Use an additional form if you participate(d) in more than 6 sports			
Type of sport(s) you have participated in (please name)	**Main sport 1**	**Other sport 2**	**Other sport 3**
Current or past participation	Current ☐ Past ☐	Current ☐ Past ☐	Current ☐ Past ☐
Year started participation			
Number of years involved in the sport			
Years in competitive sport			
Professional or amateur			
Hours of training per week (last 3 months)			
Hours of training per week (3-12 months)			
Hours of training per week (12-24 months)			

Type of sport(s) you have participated in (please name)	**Other sport 4**	**Other sport 5**	**Other sport 6**
Current or past participation	Current ☐ Past ☐	Current ☐ Past ☐	Current ☐ Past ☐
Year started participation			
Years involved in the sport			
Years in competitive sport			
Professional or amateur			
Hours of training per week (last 3 months)			
Hours of training per week (3-12 months)			
Hours of training per week (12-24 months)			

Subject No: _____

Section C. Flexibility training history	
Do you perform flexibility training (regular stretching exercises)?	Yes ☐ No ☐
If **YES**, please complete the rest of the flexibility training history section below:- If **NO**, continue completing the questionnaire from section D.	
On average, how many days a week do you perform a stretching session?	days/week
On average, how times a day do you perform a stretching session?	times/day
Please tick which muscle groups do you include in your stretching session?	☐ Hamstrings ☐ Quadriceps ☐ Calf (gastrocnemius) ☐ Calf (soleus) ☐ Groin (inner thigh) ☐ Upper body limbs ☐ Other: _____
Please tick when you stretch? (before, during and/or after exercising. You can tick more than one box)	☐ Before Exercise ☐ During Exercise ☐ After Exercise
When you stretch an individual muscle group, on average, **how long do you hold the stretch** for?	seconds
When you stretch an individual muscle group, on average, **how many times do you stretch the muscle for**?	☐ Once ☐ Twice ☐ 3 times ☐ 4 times ☐ 5 times ☐ 6 or more times

Section D. Lifestyle and habits history			
Please indicate your smoking status	Current smoker ☐	Ex smoker ☐	Never smoked ☐
If you answered yes, (past or current smoker) please complete the section on the right	Number of years of smoking:	If stopped, how many years ago:	
	What is (was) the average number of cigarettes per day:		
On average, how much alcohol do you drink per week (tots, glasses) of spirits, wine or beer?	_____ glasses beer/cider per week _____ glasses wine per week _____ tots of spirits per week		

Section E. General Personal Medical History

Do you currently suffer from any of these medical conditions:

☐ High Blood Pressure	☐ Angina/Heart Attack	☐ Asthma
☐ Emphysema	☐ Rheumatoid arthritis	☐ Osteoarthritis (wear & tear)
☐ Malignant disease (cancer)	☐ Elevated Blood Cholesterol	☐ Adrenal disorders
If Yes, what type?	☐ Diabetes mellitus	☐ Thyroid disorders
_____	☐ Renal disease	☐ Amyloidosis

Do you currently suffer from any other Connective Tissue, Rheumatological Or Muscle Diseases & Disorders?	Yes ☐ No ☐	If Yes, please select from the list below

List of some Connective Tissue and/or Rheumatic Diseases and Disorders

☐ Ankylosing Spondylitis	☐ Lipid Storage Diseases	☐ Pseudogout
☐ Aspartylglycosaminuria (AGU)	☐ Marfan Syndrome	☐ Reactive Arthritis
☐ Behcet's Syndrome	☐ Menkes Kinky Hair Syndrome	☐ Reiter's Syndrome
☐ Crohn's Disease	☐ Mucopolysaccharidoses	☐ Relapsing Polychrondritis
☐ Discoid Lupus Erythematosus	☐ Myopathies and Dystrophies	☐ Scleroderma
☐ Ehlers-Danlos syndrome (EDS)	☐ Ochronosis (Homocystinuria)	☐ Sjogren's Syndrome
☐ Eosinophilic Fascitis	☐ Osteogenesis imperfecta (OI)	☐ Systemic Lupus Erythematosus (SLE)
☐ Giant Cell (Temporal) Arthritis	☐ Polyarteritis Nodosa	☐ Systemic Sclerosis
☐ Gout	☐ Polymyalgia Rheumatica	☐ Wegener's Granulomatosis
☐ Hypersentive Vasulatis	☐ Polymyositis & Dermatomyositis	☐ Rhabdomyolysis
☐ Muscular dystrophy	☐ Myopathy	☐ Other_____

What surgical operations have you had? (please list and give dates)	Operation	Date

If female:

At what age did you start menstruating? (years)	
Are you currently using any type of contraception?	☐ Yes ☐ No
If Yes, what type of contraception are you using?	☐ Pill ☐ Injection ☐ IUD
Are you currently?	☐ Pre-menopausal (±12 cycles per year at intervals of 23– 33 days & bleeding lasts 3-7 days) ☐ Menopausal (cycles are irregular and less frequent) ☐ Post-menopausal (no longer menstruating)

Subject No: _____

Section F. Family Medical History		
Have any of your blood (biological) relatives ever had the following? **Please tick yes or no. If yes, please tick the relationship of that person to you (You may tick more than one of the relationship blocks).**		
Description		**If Yes, please indicate the relationship**
Chronic **Achilles** tendon injury	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Achilles tendon rupture	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Any other (not Achilles) tendon injury/rupture	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Any ligament injury	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Exercise associated muscle cramps	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Night muscle cramps	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Do any other members of your family suffer from elevated blood cholesterol?	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Is there any history of arthritis in your family?	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Heart Disease	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Diabetes	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother

Section G. History of Medication Use			
What medication, if any, are you currently using? (please list)	Name of medication		Years taken
Have you ever used oral corticosteroids (cortisone tablets)? (If **yes**, how long ago?)	Yes ☐ No ☐	☐ 3 months ☐ 12 months	☐ 6 months ☐ 24 or more months
Have you ever been given an injection with corticosteroids? (If **yes**, how long ago?)	Yes ☐ No ☐	☐ 3 months ☐ 12 months	☐ 6 months ☐ 24 or more months
Have you ever used fluoroquinoline antibiotics? (refer to the following list)	Yes ☐ No ☐	☐ 3 months ☐ 12 months	☐ 6 months ☐ 24 or more months

List of some fluoroquinoline antibiotics (may be used in treatment of chlamydia, pneumonia, acute bronchitis, urinary tract infections, skin and soft tissue infection):		
ADCO-CIPRIN	CIPROBAY	SANDOZ CIPROFLOXACIN
AVELON	CIPROGEN	TAFLOC
BACTIDRON	CPL ALLIANCE CIPROFLOXACIN	TARIVID
CIFLOC	DYNAFLOC	TAVANIC
CIFRAN	FACTIVE	TEQUIN
CIPLA-CIPROFLOXACIN	FLOXIN	UNIQUIN
CIPLOXX	MAXAQUIN	UTIN-400
CIPRO-HEXAL	NOROXIN	ZANOCIN
	ORPIC	

Subject No: _____

Section H. Muscle Cramping	
Have you **ever** in your athletic career suffered from **muscle cramping** (painful, spontaneous, sustained spasm of a muscle) during or immediately (within 6 hours) after exercise (in training or competition)	Yes ☐ No ☐
If **YES**, please complete the rest of the muscle cramping section below:- If **NO**, continue completing the questionnaire from section I.	
For how many years have you suffered from cramping?	(years)
Did you suffer from cramping during or after exercise in the **last 12 months**?	Yes ☐ No ☐
With what **type of exercise** is your cramping associated (You can tick more than one form of exercise)?	☐ Swimming ☐ Cycling ☐ Running
In the **last 10 races or training sessions**, how many times have you experienced cramping?	Races: _____/10 Training sessions: _____/10
What treatment/s have you had that **successfully relieved** an acute cramp? (can tick more than one)	☐ Stretching ☐ Resting ☐ Drinking fluid ☐ Ice application ☐ Massage ☐ Magnesium ☐ Salt (tablets or solution) ☐ Other (Specify: _____)
At **what point in the race or training run** do you usually first experience cramping?	☐ First quarter ☐ Second quarter ☐ Third quarter ☐ Fourth quarter ☐ After the race ☐ No pattern
In which **muscles** do you usually cramp (please list the muscle by the one which cramps most frequently (as 1) and the others after that (2-4)?	☐ Calves ☐ Hamstrings ☐ Quadriceps (thigh) ☐ Foot muscles ☐ Other (Specify: _____)
Have you **ever** suffered from cramping in your **whole body** (arms and legs)?	Yes ☐ No ☐
Have you **ever** been **admitted to hospital** following cramping?	Yes ☐ No ☐
Have you **ever** been **confused or in a coma** during or after a cramping episode?	Yes ☐ No ☐
Have you ever had **"dark urine"** in the 3 days following a cramping episode?	Yes ☐ No ☐
If you cramp, **how long** does the cramp usually last for (min)?	(minutes)
If you cramp, how **severe** is the cramp usually? (please tick).	☐ Mild: < 5 minutes and you are able to continue exercising ☐ Moderate: 5-15 minutes and you are able to continue exercising ☐ Severe: >15 minutes or if you have to STOP exercising

Subject No: _____

SECTION I. Past History of Skeletal Muscle Injury (Muscle Strain/Tear)

Please complete this section for each muscle injured. If you have had more than one muscle injury additional forms will be available.

Have you ever injured a muscle in the past? Yes ☐ No ☐

If **YES**, please complete the rest of Skeletal Muscle Injury section below:-

If **NO**, continue completing the questionnaire from section J.

If yes, please specify which muscle?
(You may tick more than one block, please select either L (left) or R (right))

Also indicate if you partially or completely tore the muscle.

Partial tear refers to tearing of a few muscle fibres with minor swelling, possible loss of strength and restriction of movement.

Complete tear refers to a tear extending across the whole muscle resulting in complete loss of muscle fuction (loss of strength, movement and ability to contract the muscle).

Muscle Group	Muscle (L-left, R-right)	Partial Tear L	Partial Tear R	Complete Tear L	Complete Tear R
Quadriceps	Vastus Lateralis	☐	☐	☐	☐
	Vastus Medialis	☐	☐	☐	☐
	Vastus Intermedius	☐	☐	☐	☐
	Rectus Femoris	☐	☐	☐	☐
Hamstring	Semitendinosus	☐	☐	☐	☐
	Semimembranosus	☐	☐	☐	☐
	Biceps femoris long	☐	☐	☐	☐
	Biceps femoris short	☐	☐	☐	☐
Hip adductor (groin)	Adductor longus	☐	☐	☐	☐
	Adductor magnu	☐	☐	☐	☐
	Adductor brevi	☐	☐	☐	☐
Calf	Gastrocnemius	☐	☐	☐	☐
	Plantaris	☐	☐	☐	☐
	Soleus	☐	☐	☐	☐
Chronic compartment syndrome of the lower leg	Anterior	Left ☐		Right ☐	
	Lateral	Left ☐		Right ☐	
	Superficial posterior	Left ☐		Right ☐	
	Deep posterior	Left ☐		Right ☐	
Other:.....					

How was the muscle injured? **(please also explain exactly how the injury occurred)**

- ☐ Contact with another player
- ☐ Contact with another object (e.g. equipment)
- ☐ No contact (sprinting)
- ☐ No contact (landing)
- ☐ No contact (kicking)
- ☐ No contact (falling)
- ☐ No contact (jumping)
- ☐ No contact (Other)
- ☐ Other .....

After sustaining the muscle injury approximately how many days were you off from training or competition?

Approximate number of days: .....

Subject No: _____

Approximate date of muscle injury?		
Investigation done to confirm the diagnosis	☐ Ultrasound ☐ MRI ☐ CT scan ☐ None	
To your knowledge, have any other members of your family suffered from any muscle pathology?	Yes ☐ No ☐	If Yes, please specify the family member ☐ Mother ☐ Father ☐ Sibling ☐ Son / daughter ☐ Other family member:..... Condition: Please choose muscle injury from the list above .....
What was the initial treatment (first 5 days)? **(You may tick more than one block.)**	☐ Rest ☐ Ice application ☐ Compression ☐ Elevation ☐ Immobilisation ☐ Medication (analgesics - pain killers) ☐ Medication (anti-inflammatory drugs) ☐ Other.....	
What was the final treatment? **(You may tick more than one block.)**	☐ Rehabilitation (stretching) ☐ Rehabilitation (strengthening) ☐ Rehabilitation (other) ☐ Strapping/taping ☐ Surgery ☐ Other.....	
Following this injury please indicate whether you were able to return to sports (indicate category).	☐ No return to any sport Return to sport but ... ☐ Limited to non-sprinting exercise ☐ Limited to non-jumping exercise ☐ Limited, not to same level as pre-injury ☐ Return to full participation in sport	
If you are able to recall, what were the weather and pitch conditions like at the time of injury?	☐ Wet and soft ground ☐ Dry, but soft ground ☐ Dry and firm ground ☐ Wet, but firm ground ☐ Other.....	

Associated injuries (Injuries sustained at the same time as the muscle injury)?	☐ Other muscle injury ☐ Tendon injury ☐ Ligament Injury ☐ Bone bruising ☐ Other.....

Section J. Past History of Tendon, Ligament or Joint Capsule Injury

Please complete this section for each injury. If you have had more than one past injury additional forms will be available.

Have you **ever** in your suffered from **a tendon or ligament injury** (pain, swelling, stiffness) in any tendon (including Achilles tendon, knee tendons, and shoulder tendons) or ligaments (partial or complete tear)?	Yes ☐ No ☐

If **YES**, please complete the rest of the section below:-

If **NO**, continue completing the questionnaire from section L.

Please tick which **tendon/s** you have injured? (next column on the tright)

Also indicate (tick) if your injured tendon was longsatnding pain (tendinopathy) or an acute tear/rupture

Tendon		Longstanding Pain (Tendinopathy) Left Right	Acute Tear/ Rupture Left Right
Foot and ankle:	☐ Achilles tendon	☐☐	☐☐
	☐ Tibialis posterior	☐☐	☐☐
	☐ Plantar fascia	☐☐	☐☐
Knee:	☐ Patellar tendon	☐☐	☐☐
Elbow and wrist:	☐ Wrist extensor tendon	☐☐	☐☐
Shoulder:	☐ Rotator cuff	☐☐	☐☐
Other: _____		☐☐	☐☐

Please tick which **ligament/s** you have injured? (next column on the tright)

Also indicate if your sprained or completely tore the ligamant.

Ligament	Sprain Left Right	Complete Tear Left Right
☐ Shoulder ligaments	☐☐	☐☐
☐ Elbow ligaments	☐☐	☐☐
☐ Wrist ligaments	☐☐	☐☐
☐ Finger ligaments	☐☐	☐☐
☐ Knee (ACL)	☐☐	☐☐
☐ Knee (MCL)	☐☐	☐☐
☐ Knee (PCL)	☐☐	☐☐
☐ Knee (LCL)	☐☐	☐☐
☐ Ankle lateral ligaments	☐☐	☐☐
☐ Ankle medial ligaments	☐☐	☐☐
☐ Spinal ligaments	☐☐	☐☐
☐ Other: _____	☐☐	☐☐

Have you ever suffered from any of the following joint capsule injuries?	☐ Acute shoulder dislocation ☐ Chronic shoulder instability ☐ Chronic ankle instability ☐ Other: _____

SECTION K. MEDICAL DETAILS OF TENDON INJURIES

Symptoms

How many times have you had tendon injuries? [1]Sudden onset is within a few seconds or minutes [2]Gradual onset is over days or weeks	Tendon Injured	Date of Injury	Acute or Chronic Injury	Sudden[1] or Gradual[2] Onset
	1			
	2			
	3			
	4			
	5			

Please complete a **separate form**, Part K only, for each Tendon Injury you have had			
Injury Number (1,2,3,4,or 5)	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ _____		
Which tendon did you injure?	Rotator cuff tendon ☐ • Supraspinatus ☐ • Infraspinatus ☐ • teres minor ☐	Patellar tendon ☐ Wrist extensor tendons ☐ Achilles tendon ☐	
Which side was injured?	☐ Left	☐ Right	☐ Both
Which region of your tendon was injured? Please indicate on a diagram. (Only if applicable)	☐ Upper 1/3	☐ Middle 1/3	☐ Lower 1/3
To what extent was your Tendon ruptured?	☐ Complete	☐ Partial	☐ None
How were you injured? (e.g. sport, walking)			
Grade of injury **at the time** of injury	☐ pain only after exercise ☐ pain during exercise, but did not cause you to alter training ☐ pain during exercise, which causes you to alter training ☐ pain which causes you to stop training ☐ no pain ☐ not sure ☐ Other (Specify _____)		
Grade of injury **currently**	☐ pain only after exercise ☐ pain during exercise, but did not cause you to alter training. ☐ pain during exercise, which causes you to alter training ☐ pain which causes you to stop training ☐ no pain ☐ not sure ☐ Other (Specify _____)		

Which of the following symptoms were present **before** the injury	☐ Pain (less than 1 week) ☐ Stiffness ☐ Pain (1-4 weeks) ☐ Swelling ☐ Pain (> 4 weeks) ☐ None
Which of the following symptoms were present **after** the injury	☐ Pain (less than 1 week) ☐ Stiffness ☐ Pain (1-4 weeks) ☐ Swelling ☐ Pain (> 4 weeks) ☐ None
If you have or had chronic tendon pain, what seems to alleviate the pain?	
Diagnosis	
Which type of Tendon Disease were you diagnosed with e.g. Rupture, Tendinitis, etc.	
Diagnosed by (Please indicate the name and contact number of the clinician who diagnosed you)	☐ Doctor _____ ☐ Physiotherapist _____ ☐ Biokineticist _____ ☐ Podiatrist _____ ☐ Other _____
If you had a tendon rupture. How was it treated?	☐ Surgically ☐Non-surgically
If applicable, who was the surgeon?	Surgeon _____Phone _____
If applicable, what diagnostic imaging was performed?	☐ Ultrasound ☐ MRI ☐ CT Other _____
If applicable, who did the imaging?	Clinician _____Phone_____

Subject No: _____

Section L. Details of Any Other Chronic (Longstanding) Current Injury

Please complete this section for each injury. If you have had more than one past injury additional forms will be available.

What was the approximate date when you first became aware of the injury?	Month Year
Please indicate which side of your body is injured (if applicable)	☐ Right ☐ Left
Please indicate which anatomical area is currently injured	☐ Head ☐ Neck ☐ Face ☐ Front chest ☐ Back chest ☐ Shoulder ☐ Upper arm ☐ Elbow ☐ Forearm ☐ Wrist ☐ Finger ☐ Lower back ☐ Hip ☐ Thigh ☐ Hamstring ☐ Quadriceps ☐ Knee ☐ Shin ☐ Achilles ☐ Ankle ☐ Foot Other (Specify: _____)
Please indicate the type of structure that was injured	☐ Muscle ☐ Tendon ☐ Bone ☐ Ligament ☐ Joint Other (Specify): _____)
Please indicate in which sport (discipline) the injury occurred	☐ Running ☐ Soccer ☐ Rugby ☐ Cricket ☐ Hockey Other (Specify: _____)
Please indicate the severity of the injury (tick one box please)	☐ I only experience symptoms after exercise - Grade 1 ☐ I experience symptoms during exercise, but it does not interfere with exercise - Grade 2 ☐ I experience symptoms during exercise that may interfere with my training/ competition - Grade 3 ☐ I am so painful that I may not be able to train or compete - Grade 4
Please indicate how your injury was treated to date (you can tick more than one)?	☐ Rest ☐ Stretches ☐ Physiotherapy ☐ Surgery ☐ Strengthening exercises ☐ Equipment change ☐ Tablets ☐ Cortisone injection ☐ Other injection ☐ Orthotics Other (Specify: _____)

5. Diagnostic criteria forms

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: + 27 21 650 4561
Fax: + 27 21 686 7530

CLINICAL DIAGNOSIS OF ACHILLES TENDINOPATHY

SUBJECT NUMBER/CODE: _____

Clinical criteria [1,2]	Present
Gradual progressive pain over the posterior lower leg - Achilles tendon area (> 6 weeks)	
Early morning pain	
Early morning stiffness	
History of swelling over the Achilles tendon area	
Tenderness to palpation over the Achilles tendon	
Palpable nodular thickening over the affected Achilles	
Positive "shift" test (movement of the nodular area with plantar- /dorsi-flexion)	

Other criteria	Present
Confirmation of the diagnosis by ultrasound *	
Confirmation of the diagnosis by MRI *	
Confirmation of the diagnosis by CT scan *	

*: One of these criteria must be present to confirm the diagnosis

Date: _____ / _____ / _____

Investigator: Prof M Schwellnus

Signature: _____

References:

1. Schepsis AA, Jones H, Haas AL. Achilles tendon disorders in athletes. *Am.J Sports Med* 2002;**30**:287-305.
2. Kader D, Saxena A, Movin T, Maffulli N. Achilles tendinopathy: some aspects of basic science and clinical management. *Br.J Sports Med* 2002;**36**:239-49.

(B) ALLELIC DISCRIMINATION METHODOLOGY

1. PCR Conditions for COL1A1 rs1800012 G/T

Primer pairs

Primary reaction	
Forward Primer	5'-GGA AGA CCC GGG TTA TTG CT-3'
Reverse Primer	5'-CGC TGA AGC CAA GTG AAA TA-3'
Secondary reaction	
Forward Primer	5'-TAA CTT CTG GAC TAT TTG CGG ACT TTT TGG-3'
Reverse Primer	5'-GTC CAG CCC TCA TCC TGG CC-3'

The secondary reverse primer was designed to contain two mutated nucleotides (underlined in the primer sequence) which introduced a restriction site (TGG'CCA) for the restriction endonuclease *Msc*I at the 3' end of the 260 bp secondary PCR product when amplifying the T allele.

PCR Conditions

Polymerase	Taq DNA Polymerase
Primary Cycling	5 cycles
Primary Annealing temperature	70^{o}C
Secondary Cycling	27 cycles
Secondary Annealing temperature	58^{o}C
Mg2+ Concentration	2.0 mM
Amplicon size	260 bp

Restriction Conditions

Nuclease	*Msc*I
Cutting site	TGG'CCA
Incubation temperature	37^{o}C
Fragment sizes	G allele – 260 bp T allele – 242 bp, 18 bp

2. PCR Conditions for COL5A1 rs12722 T/C

Primer pairs

Forward Primer	5'-GAA GAC GTT TCT GGA GGA TC-3'
Reverse Primer	5'-GGA GGC ACC TGC AGA ATG AC-3'

PCR Conditions

Polymerase	Taq DNA Polymerase
Cycling	35 cycles
Annealing temperature	53^{o}C
Mg2+ Concentration	1.5 mM
Amplicon size	667 bp

Restriction Conditions

Nuclease	*Bstu*I
Cutting site	CG'CG
Incubation temperature	60^{o}C
Fragment sizes	T allele – 351 bp, 316 bp C allele – 316 bp, 271 bp, 80 bp

3. PCR Conditions for COL12A1 rs970547 A/G

Primer pairs

Forward Primer	5'-GAG AAT CCA GAA CAG CTC CAC CAG-3'
Reverse Primer	5'-CAT GGC TAG TAT GGG ACA G-3'

The *COL12A1* forward primer was designed to contain a mutated nucleotide (underlined in the primer sequence) which introduces and additional restriction site (AG'CT) for the A*lu*I restriction endonuclease.

PCR Conditions

Polymerase	Taq DNA Polymerase
Cycling	30 cycles
Annealing temperature	58^{o}C
Mg2+ Concentration	2.0 mM
Amplicon size	615 bp

Restriction Conditions

Nuclease	*Alu*I
Cutting site	AG'CT
Incubation temperature	37^{o}C
Fragment sizes	G allele – 599 bp, 16 bp A allele – 460 bp, 139 bp, 16 bp

4. PCR Conditions for COL3A1 rs1800255 G/A

PCR Conditions

Taqman SNP Genotyping Assay (Applied Biosystems, Foster City, CA, USA).

Assay ID: AHGJQCQ

Primer pairs

Forward Primer	5'- CGT GGA CCT CCT GGA TTG G -3'
Reverse Primer	5'- GAA TGC TGT GGA GTT ACC TTT CCT -3'

Probe sequences

Allele 1 (G allele)	5'- TAG AGG TGG AGC TGG TC -3'
Allele 2 (A allele)	5'- TAG AGG TGG AAC TGG TC -3'

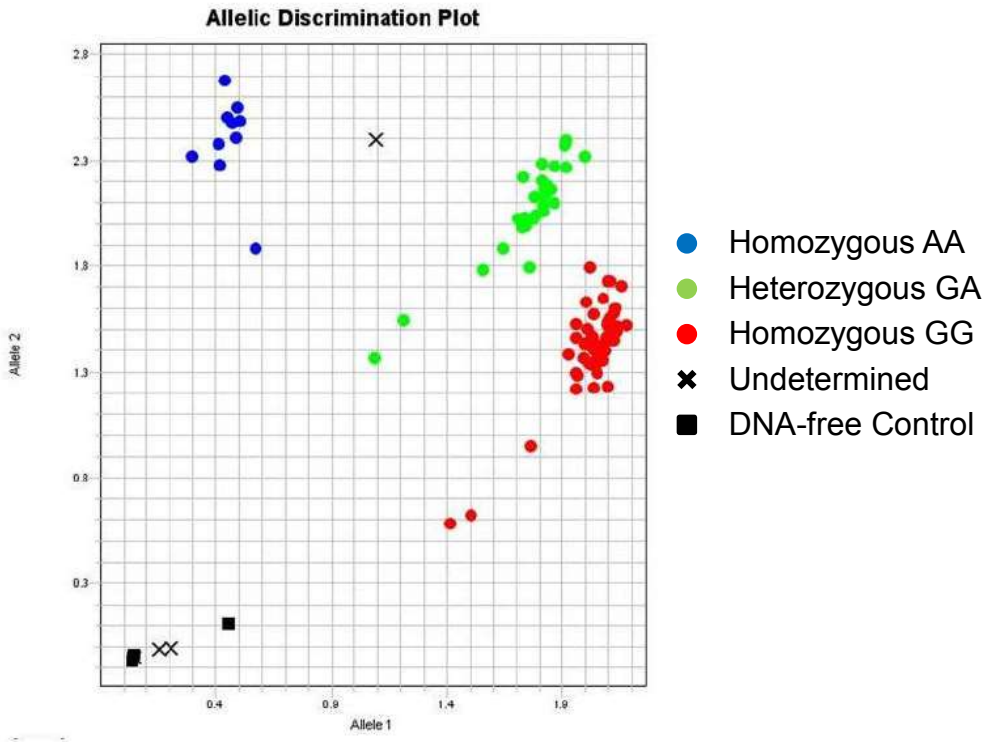


Figure B1. A typical allelic discrimination plot for rs1800255 using the Taqman SNP Genotyping assay described above.

5. PCR Conditions for COL6A1 rs35796750 T/C

PCR Conditions

Taqman SNP Genotyping Assay (Applied Biosystems, Foster City, CA, USA).

Assay ID: AH89JPA

Primer pairs

Forward Primer	5'- GGC CTT GTC CCC AGA AAG AC -3'
Reverse Primer	5'- CCA CGG AGA CCA CCT GTG -3'

Probe sequences

Allele 1 (C allele)	5'- TGT GGC GCA GCC TG -3'
Allele 2 (T allele)	5'- TGT GGC ACA GCC TG -3'

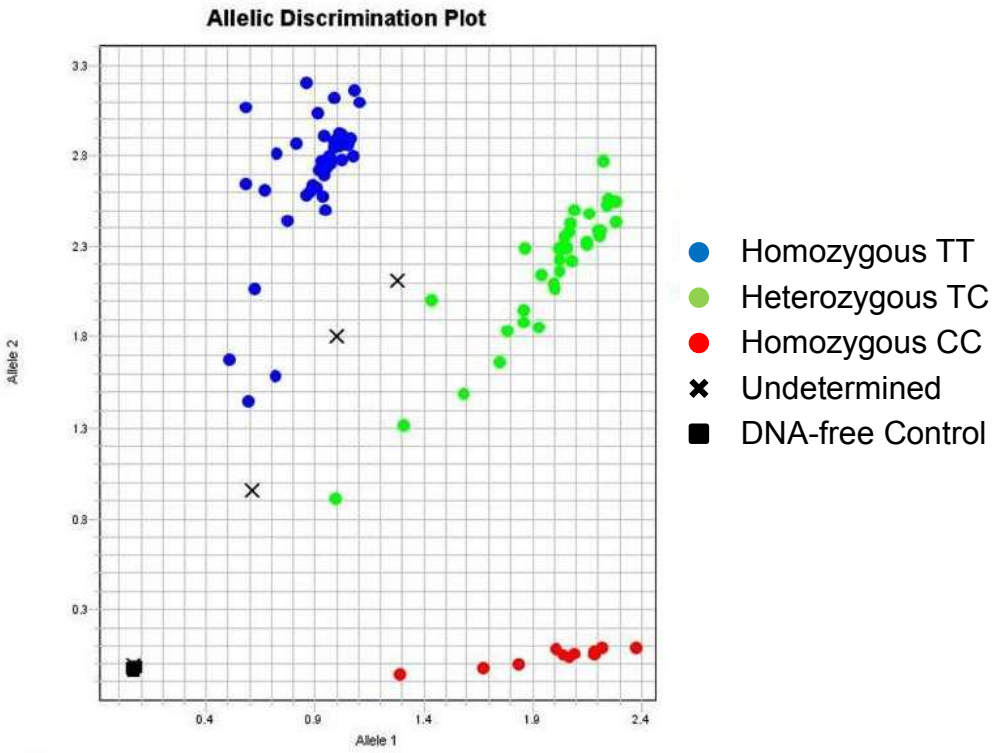


Figure B2. A typical allelic discrimination plot for rs35796750 using the Taqman SNP Genotyping assay described above.

(C) SUPPLEMENTARY RESULTS

1. Participant and family history of soft tissue injuries for the male and female participants.

Male Participants	CON (145)	ACL (177)	p-value[a]	NON (103)	p-value[b]
Participant previous ligament injury	40.0 (58)	44.1 (78)	0.462	41.7 (43)	0.076
• **Knee ligament injury[c]**	4.1 (6)	11.9 (21)	**0.010**	8.7 (9)	0.134
• **Ankle ligament strain[d]**	21.4 (31)	28.2 (50)	0.158	30.1 (31)	0.118
Participant previous tendon injury	20.7 (2)	23.7 (42)	0.515	25.2 (26)	0.398
Participant history of joint capsule disease	11.0 (16)	8.5 (15)	0.439	9.7 (10)	0.737
Family history ligament injury	26.2 (38)	37.3 (66)	**0.034**	36.9 (38)	0.072
• **ACL injury**	2.1 (3)	15.3 (27)	**<0.001**	15.5 (16)	**<0.001**
Family history tendon injury	6.2 (9)	10.7 (19)	0.151	8.7 (9)	0.582
Female	**CON (90)**	**ACL (65)**	**p-value[a]**	**NON (33)**	**p-value[b]**
Participant previous ligament injury	24.4 (22)	43.1 (28)	**0.014**	48.5 (16)	**0.011**
• **Knee ligament injury[c]**	1.1 (1)	10.8 (7)	**0.007**	18.2 (6)	**<0.001**
• **Ankle ligament strain[d]**	22.2 (20)	30.8 (20)	0.230	36.4 (12)	0.113
Participant previous tendon injury	18.9 (17)	15.4 (10)	0.570	27.3 (9)	0.313
Participant history of joint capsule disease	7.8 (7)	15.4 (10)	0.135	24.2 (8)	**0.013**
Family history ligament injury	27.8 (25)	44.6 (29)	**0.030**	54.5 (18)	**0.006**
• **ACL injury**	2.2 (2)	20.0 (13)	**<0.001**	33.3 (11)	**<0.001**
Family history tendon injury	11.1 (10)	13.8 (9)	0.608	12.1 (4)	0.876

Values are expressed as percentages with the number of participants (n) indicated in parentheses. Values in bold typeset are significant ($p<0.05$). CON, apparently healthy controls. ACL, anterior cruciate ligament. NON, non-contact mechanism of injury.

[a] CON vs. ACL

[b] CON vs. NON

[c] includes the posterior cruciate ligament, the lateral collateral ligament and the medial collateral ligament.

[d] includes the lateral and medial ankle ligaments.

2. Genotype frequency distributions when participants with a self-reported history of any previous ligament injury were excluded from the analysis.

	***COL3A1* rs1800255 Genotype**			**n**	**p-value**
	GG	**GA**	**AA**		
All Participants					
CON	53.5 (77)	37.5 (54)	9.0 (13)	144	
ACL	56.6 (69)	36.9 (45)	6.6 (8)	122	0.729[a]
NON	55.9 (38)	36.8 (25)	7.4 (5)	68	0.901[b]
Male Participants					
CON	54.6 (42)	36.4 (28)	9.1 (7)	77	
ACL	60.4 (55)	36.3 (33)	3.3 (3)	91	0.272[a]
NON	59.3 (32)	35.2 (19)	5.6 (3)	54	0.721[b]
Female Participants					
CON	52.2 (35)	38.8 (26)	9.0 (6)	67	
ACL	45.2 (14)	38.7 (12)	16.1 (5)	31	nd
NON	42.9 (6)	42.9 (6)	14.3 (2)	14	nd

Values are expressed as percentages with the number of participants indicated in parentheses. The total number (n) of participants genotyped in each individual or combined cohort is also indicated. Values in bold typeset are significant ($p<0.05$). All the ACL ruptures were diagnosed at surgery.

nd, not determined.

[a] CON vs. ACL

[b] CON vs. NON

	COL6A1 rs35796750 Genotype			n	p-value
	TT	TC	CC		
All Participants					
CON	34.9 (44)	46.0 (58)	19.1 (24)	126	
ACL subset 1	33.3 (19)	54.4 (31)	12.3 (7)	57	0.438[a]
NON subset 1	31.3 (10)	56.3 (18)	12.5 (4)	32	nd
Male Participants					
CON	37.5 (27)	41.7 (30)	20.8 (15)	72	
ACL subset 1	32.5 (13)	55.0 (22)	12.5 (5)	40	0.339
NON subset 1	26.1 (6)	56.5 (13)	17.4 (4)	23	nd
Female Participants					
CON	31.5 (17)	51.9 (28)	16.7 (9)	54	
ACL subset 1	35.3 (6)	52.9 (9)	11.8 (2)	17	nd
NON subset 1	44.4 (4)	55.6 (5)	0.0 (0)	9	nd

Values are expressed as percentages with the number of participants indicated in parentheses. The total number (n) of participants genotyped in each individual or combined cohort is also indicated. Values in bold typeset are significant ($p<0.05$). All the ACL ruptures were diagnosed at surgery.

nd, not determined.

[a] CON vs. ACL

[b] CON vs. NON

	COL5A1 rs12722 Genotype			n	p-value
	TT	**TC**	**CC**		
Male Participants					
CON	31.4 (27)	54.7 (47)	14.0 (12)	86	
ACL	28.6 (26)	59.3 (54)	12.1 (11)	91	0.816[a]
NON	29.6 (16)	59.3 (32)	11.1 (6)	54	0.832[b]
Female Participants					
CON	29.4 (20)	41.2 (28)	29.4 (20)	68	
ACL	33.3 (11)	57.6 (16)	9.1 (3)	33	nd
NON	23.5 (4)	58.8 (10)	17.7 (3)	17	nd
	COL12A1 rs970547 Genotype			n	p-value
	AA	**AG**	**GG**		
Male Participants					
CON	59.5 (50)	36.9 (31)	3.6 (3)	84	
ACL	58.6 (51)	34.5 (30)	6.9 (6)	87	0.614[a]
NON	58.8 (30)	31.4 (16)	9.8 (5)	51	0.307[b]
Female Participants					
CON	60.6 (40)	36.4 (24)	3.0 (2)	66	
ACL	60.6 (20)	27.3 (9)	12.1 (4)	33	nd
NON	52.9 (9)	29.4 (5)	17.7 (3)	17	nd

Values are expressed as percentages with the number of participants indicated in parentheses. The total number (n) of participants genotyped in each individual or combined cohort is also indicated. Values in bold typeset are significant ($p<0.05$). All the ACL ruptures were diagnosed at surgery.

nd, not determined.

[a] CON vs. ACL

[b] CON vs. NON

3. A List of Reported Contact and Non-contact non-jumping and Non-contact jumping Sports, as well as Skiing

Contact Sports:

Boxing

Gaelic Football

Hurling

Judo

Jujitsu

Karate

Rugby

Soccer

Touch Rugby

Non-contact non-jumping Sports:

Canoeing

Cricket

Cycling

Equestrian

Golf

Hockey

Horse riding

Javelin

Kayaking

Mountain Biking

Rock climbing

Rowing

Running

Sailing

Spinning

Squash

Swimming

Tennis

Yoga

Non-contact jumping Sports:

Athletics

Ballet

Basketball

BMX

Dancing

Gymnastics

Ice skating

Kite surfing

Motorcross

Netball

Paragliding

Skateboarding

Volleyball

Yachting

Skiing Sports:

Water skiing

Snowboarding

Snow skiing

Surf ski

Wakeboarding

4. COL3A1 rs1800255 and Achilles tendinopathy

A comparison of the genotype and allele frequency distributions for *COL3A1* rs1800255 (G/A) between the South African and Australian Achilles tendinopathy groups and their respective control groups. Genotype and allele frequency distributions after combining the cohorts are also reported. Data taken from Saunders (2013) [170].

	***COL3A1* rs1800255 Genotype**			**n**	**Genotype p-value**	**Minor Allele**	**Allele p-value**
	GG	**GA**	**AA**				
SA TEN	56.7 (51)	35.6 (32)	7.8 (7)	90		25.6 (46)	
SA CON	54.9 (89)	34.6 (56)	10.5 (17)	162	0.781	27.8 (90)	0.590
AUS TEN	63.3 (50)	30.4 (24)	6.3 (5)	79		21.5 (34)	
AUS CON	56.2 (109)	36.1 (70)	7.7 (15)	194	0.558	25.8 (100)	0.295
All TEN	59.8 (101)	33.1 (56)	7.1 (12)	169		23.7 (80)	
All CON	55.6 (198)	35.4 (126)	9.0 (32)	356	0.606	26.7 (190)	0.296

Values are expressed as percentages with the number of participants indicated in parentheses. The total number (n) of participants genotyped in each individual or combined cohort is also indicated.

SA, South African; AUS, Australia; All, South Africa and Australia; TEN, participants with diagnosed Achilles tendinopathy; CON, apparently healthy control participants.

5. COL5A1 rs12722 and Achilles tendinopathy

A comparison of the genotype and allele frequency distributions for *COL5A1* rs12722 (T/C) between the South African and Australian Achilles tendinopathy groups and their respective control groups. Genotype and allele frequency distributions after combining the cohorts are also reported. Data taken from Mokone et al. and September et al. [131;179].

	***COL5A1* rs12722 Genotype**			**n**	**Genotype p-value**	**Minor Allele**	**Allele p-value**
	TT	**TC**	**CC**				
All Participants							
SA TEN	36.6 (34)	50.5 (47)	12.9 (12)	93		38.2 (71)	
SA CON	29.6 (47)	43.4 (69)	27.0 (43)	159	**0.031**	48.7 (155)	**0.021**
AUS TEN	20.0 (17)	68.2 (58)	11.8 (10)	85		45.9 (78)	
AUS CON	35.2 (74)	40.5 (85)	24.3 (51)	210	**<0.001**	44.5 (187)	0.764
All TEN	28.7 (51)	59.0 (105)	12.4 (22)	178		41.9 (149)	
All CON	32.8 (121)	41.7 (154)	25.5 (94)	369	**<0.001**	46.3 (342)	0.162

Values are expressed as percentages with the number of participants indicated in parentheses. The total number (n) of participants genotyped in each individual or combined cohort is also indicated. Significant p-values are indicated in bold.

SA, South African; AUS, Australia; All, South Africa and Australia; TEN, participants with diagnosed Achilles tendinopathy; CON, apparently healthy control participants.

6. COL12A1 rs970547 and Achilles tendinopathy

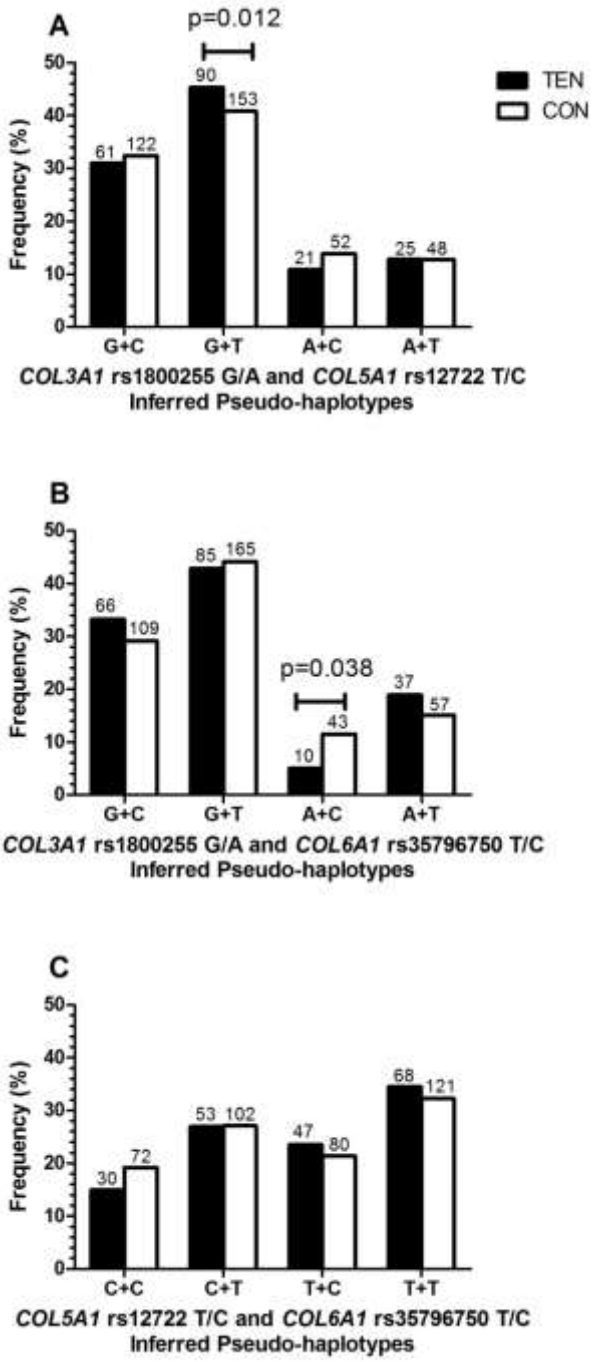
A comparison of the genotype and allele frequency distributions for *COL12A1* rs970547 (A/G) between the South African Achilles tendinopathy and control groups. Data taken from September et al. [181].

	***COL12A1* rs970547 Genotype**			**n**	**Genotype p-value**	**Minor Allele**	**Allele p-value**
	AA	**AG**	**GG**				
SA TEN	44.7 (38)	47.1 (40)	8.2 (7)	85		31.8 (54)	
SA CON	45.3 (72)	47.8 (76)	6.9 (11)	159	0.932	30.8 (98)	0.838

Values are expressed as percentages with the number of participants indicated in parentheses. The total number (n) of participants genotyped in each individual or combined cohort is also indicated.

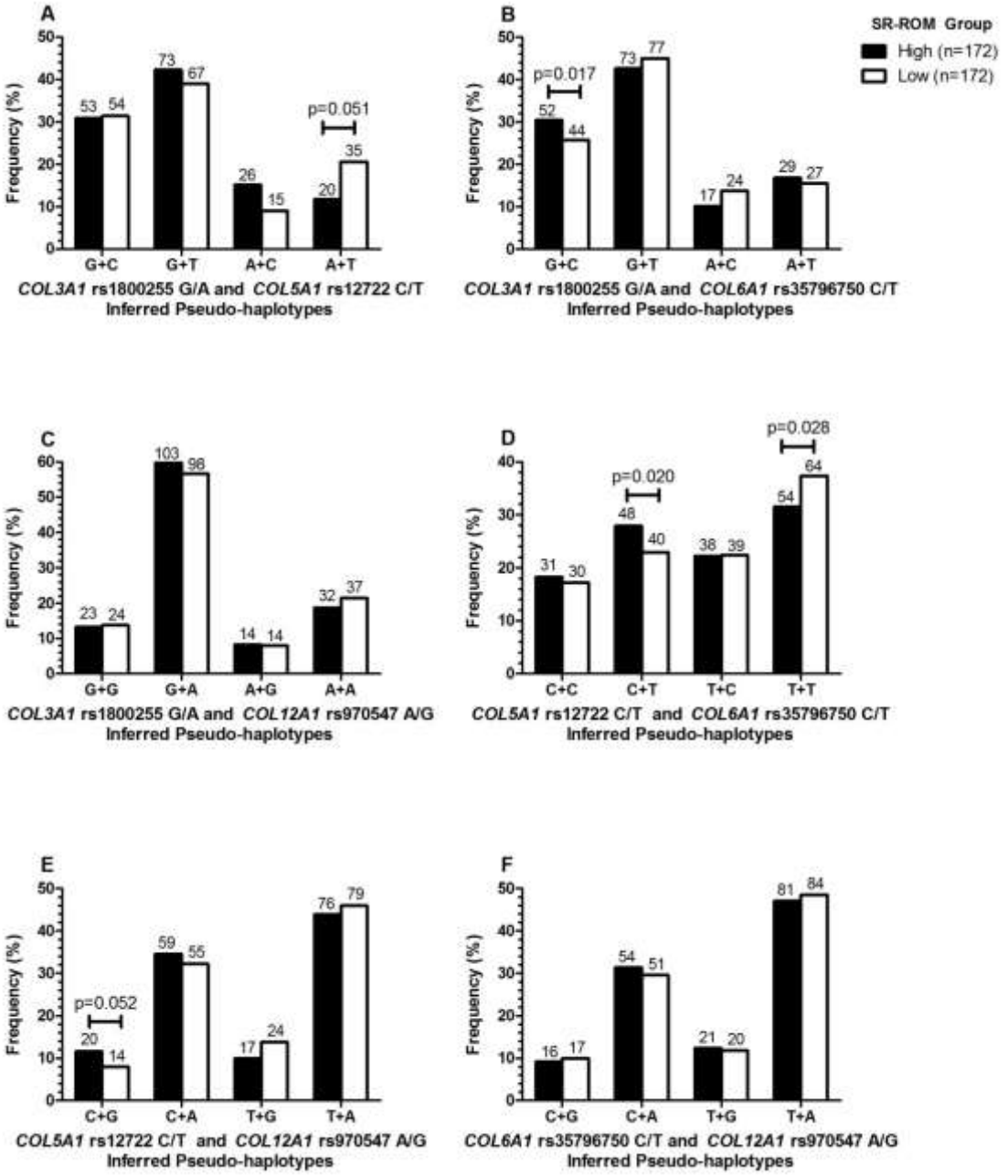
SA, South African; TEN, participants with diagnosed Achilles tendinopathy; CON, apparently healthy control participants.

7. Two-gene inferred pseudo-haplotypes constructed from COL3A1 rs1800255, COL5A1 rs12722 and/or COL6A1 rs35796750 and Achilles tendinopathy.



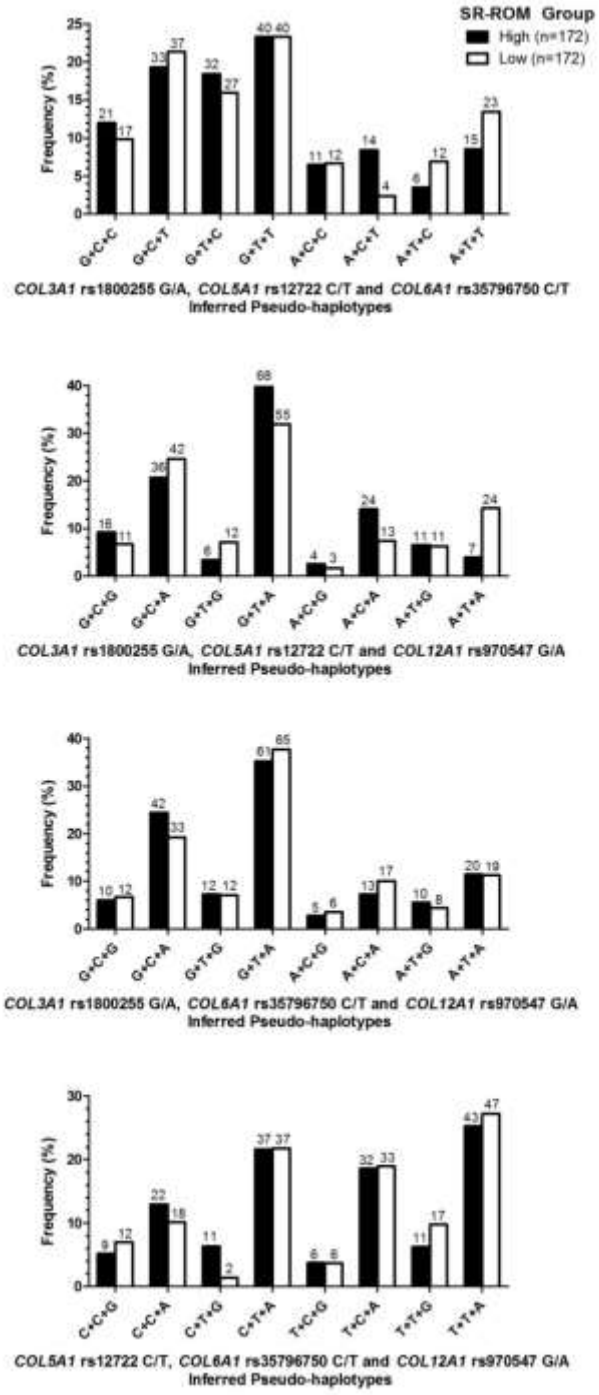
Inferred pseudo-haplotype frequencies, constructed from *COL3A1* rs1800255, *COL5A1* rs12722 and/or *COL6A1* rs37596750, in South African and Australian participants with clinically diagnosed Achilles tendinopathy (TEN group) and apparently healthy controls (CON group). Numbers of participants are listed above each column.

8. Two-gene inferred pseudo-haplotypes constructed from COL3A1 rs1800255, COL5A1 rs12722, COL6A1 rs35796750 and/or COL12A1 rs970547 and Range of Motion.



Inferred pseudo-haplotype frequencies, constructed from *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs37596750 and/or *COL12A1* rs970547, between the High and Low SR-ROM (sit-and-reach ROM) groups. Numbers of participants are listed above each column.

9. Three-gene inferred pseudo-haplotypes constructed from COL3A1 rs1800255, COL5A1 rs12722, COL6A1 rs35796750 and/or COL12A1 rs970547 and Range of Motion.



Inferred pseudo-haplotype frequencies, constructed from *COL3A1* rs1800255, *COL5A1* rs12722, *COL6A1* rs37596750 and/or *COL12A1* rs970547, between the High and Low SR-ROM (sit-and-reach ROM) groups. Numbers of participants are listed above each column.

10. Genotype Frequency Distributions for COL3A1 rs1800255, COL5A1 rs12722, COL6A1 rs35796750 and COL12A1 rs970547 between the participants that had never played rugby and the recreational rugby players.

	***COL3A1* rs1800255 Genotype**			**p-value**
	GG	**GA**	**AA**	
NR CON	54.2 (39)	38.9 (28)	6.9 (5)	0.988
Recreational	52.9 (37)	40.0 (28)	7.1 (5)	
	***COL5A1* rs12722 Genotype**			**p-value**
	TT	**TC**	**CC**	
NR CON	27.6 (21)	51.3 (39)	21.1 (16)	0.152
Recreational	41.0 (32)	46.2 (36)	12.8 (10)	
	***COL6A1* rs35796750 Genotype**			**p-value**
	TT	**TC**	**CC**	
NR CON	35.1 (26)	52.7 (39)	12.2 (9)	0.469
Recreational	32.5 (25)	48.1 (37)	19.5 (15)	
	***COL12A1* rs970547 Genotype**			**p-value**
	AA	**AG**	**GG**	
NR CON	57.7 (45)	39.7 (31)	2.6 (2)	0.211
Recreational	52.6 (40)	38.2 (29)	9.2 (7)	

Values are expressed as percentages. Numbers of participants are indicated in parentheses.

NR CON, never played rugby before.

(D) Quantitative Real-Time PCR Methodology

Primer pairs

***COL6A1* primers**	
Forward Primer	5’- TCA GAA TAG TGA TGT GTT CGA CGT T -3’
Reverse Primer	5’- AGC AAC ATG GAT ATG GTT CAG AAA -3’
***COL1A1* primers**	
Forward Primer	5’- AGC CAG CAG ATC GAG AAC A -3’
Reverse Primer	5’- TCT TGT CCT TGG GGT TCT T -3’
***β-Actin* primers**	
Forward Primer	5’- CCT CGC CTT TGC CGA TCC G -3’
Reverse Primer	5’- GCC GGA GCC GTT GTC GAC G -3’

PCR Conditions

Cycling	40 cycles
Annealing temperature	60^{o}C

REFERENCES

(1) Abrahams S, Posthumus M and Collins M. A polymorphism within a functional region of the COL5A1 gene is independently associated with ultra-endurance running performance and joint range of motion. *Int J Sports Physiol Perform*. 2013;9(3):583-90.

(2) Abrahams Y, Laguette MJ, Prince S, et al. Polymorphisms within the COL5A1 3'-UTR that alters mRNA structure and the MIR608 gene are associated with achilles tendinopathy. *Ann Hum Genet*. 2013;77(3):204-14.

(3) Ahmetov II, Mozhayskaya IA, Flavell DM, et al. PPARalpha gene variation and physical performance in Russian athletes. *Eur J Appl Physiol*. 2006;97(1):103-8.

(4) Altshuler D, Daly MJ and Lander ES. Genetic mapping in human disease. *Science*. 2008;322(5903):881-8.

(5) Amenta PS, Gil J and Martinez-Hernandez A. Connective tissue of rat lung. II: Ultrastructural localization of collagen types III, IV, and VI. *J Histochem Cytochem*. 1988;36(9):1167-73.

(6) Anderson B and Burke ER. Scientific, medical, and practical aspects of stretching. *Clin Sports Med*. 1991;10(1):63-86.

(7) Archambault JM, Wiley JP and Bray RC. Exercise loading of tendons and the development of overuse injuries. A review of current literature. *Sports Med*. 1995;20(2):77-89.

(8) Attwooll C, Lazzerini DE and Helin K. The E2F family: specific functions and overlapping interests. *EMBO J*. 2004;23(24):4709-16.

(9) Banos CC, Thomas AH and Kuo CK. Collagen fibrillogenesis in tendon development: current models and regulation of fibril assembly. *Birth Defects Res C Embryo Today*. 2008;84(3):228-44.

(10) Barnes CJ, Van Steyn SJ and Fischer RA. The effects of age, sex, and shoulder dominance on range of motion of the shoulder. *J Shoulder Elbow Surg*. 2001;10(3):242-6.

(11) Battie MC, Levalahti E, Videman T, et al. Heritability of lumbar flexibility and the role of disc degeneration and body weight. *J Appl Physiol*. 2008;104(2):379-85.

(12) Baumgarten I. A comparison of metabolic pathway dynamics in man and other mammals. *Masters Diploma in Medical Technology, Cape Technikon, Cape Town, South Africa*. 1993.

(13) Bell RD, Shultz SJ, Wideman L, et al. Collagen gene variants previously associated with anterior cruciate ligament injury risk are also associated with joint laxity. *Sports Health*. 2012;4(4):312-8.

(14) Bell W, Colley JP, Evans WD, et al. ACTN3 genotypes of Rugby Union players: distribution, power output and body composition. *Ann Hum Biol*. 2012;39(1):19-27.

(15) Bell W, Colley JP, Gwynne JR, et al. ACE ID genotype and leg power in Rugby Union players. *J Sports Med Phys Fitness*. 2010;50(3):350-5.

(16) Bernard M, Yoshioka H, Rodriguez E, et al. Cloning and sequencing of pro-alpha 1 (XI) collagen cDNA demonstrates that type XI belongs to the fibrillar class of collagens and reveals that the expression of the gene is not restricted to cartilagenous tissue. *J Biol Chem*. 1988;263(32):17159-66.

(17) Bertini E and Pepe G. Collagen type VI and related disorders: Bethlem myopathy and Ullrich scleroatonic muscular dystrophy. *Eur J Paediatr Neurol*. 2002;6(4):193-8.

(18) Birk DE. Type V collagen: heterotypic type I/V collagen interactions in the regulation of fibril assembly. *Micron*. 2001;32(3):223-37.

(19) Birk DE, Fitch JM, Babiarz JP, et al. Collagen fibrillogenesis in vitro: interaction of types I and V collagen regulates fibril diameter. *Journal of Cell Science*. 1990;95 (Pt 4):649-57.

(20) Boden BP, Dean GS, Feagin JA, Jr., et al. Mechanisms of anterior cruciate ligament injury. *Orthopedics*. 2000;23(6):573-8.

(21) Bonaldo P, Braghetta P, Zanetti M, et al. Collagen VI deficiency induces early onset myopathy in the mouse: an animal model for Bethlem myopathy. *Hum Mol Genet*. 1998;7(13):2135-40.

(22) Bonaldo P, Braghetta P, Zanetti M, et al. Collagen VI deficiency induces early onset myopathy in the mouse: an animal model for Bethlem myopathy. *Hum Mol Genet*. 1998;7(13):2135-40.

(23) Boot-Handford RP, Tuckwell DS, Plumb DA, et al. A novel and highly conserved collagen (pro(alpha)1(XXVII)) with a unique expression pattern and unusual molecular characteristics establishes a new clade within the vertebrate fibrillar collagen family. *J Biol Chem*. 2003;278(33):31067-77.

(24) Borsa PA, Dover GC, Wilk KE, et al. Glenohumeral range of motion and stiffness in professional baseball pitchers. *Med Sci Sports Exerc*. 2006;38(1):21-6.

(25) Boudko SP, Zientek KD, Vance J, et al. The NC2 domain of collagen IX provides chain selection and heterotrimerization. *J Biol Chem*. 2010;285(31):23721-31.

(26) Bray MS, Hagberg JM, Perusse L, et al. The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update. *Med Sci Sports Exerc*. 2009;41(1):35-73.

(27) Brooks JH, Fuller CW, Kemp SP, et al. Epidemiology of injuries in English professional rugby union: part 1 match injuries. *Br J Sports Med*. 2005;39(10):757-66.

(28) Brown JC, Miller CJ, Posthumus M, et al. The COL5A1 Gene, Ultra-Marathon Running Performance, and Range of Motion. *Int J Sports Physiol and Perf*. 2011;6(4):485-96.

(29) Brown JC, Miller CJ, Schwellnus MP, et al. Range of motion measurements diverge with increasing age for COL5A1 genotypes. *Scand J Med Sci Sports*. 2011;21(6):e266-e272.

(30) Bryne JC, Valen E, Tang MH, et al. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res*. 2008;36(Database issue):D102-D106.

(31) Burch GH, Gong Y, Liu W, et al. Tenascin-X deficiency is associated with Ehlers-Danlos syndrome. *Nat Genet*. 1997;17(1):104-8.

(32) Burger M, de WH and Collins M. The COL5A1 gene is associated with increased risk of carpal tunnel syndrome. *Clin Rheumatol*. 2014;10.1007/s10067-014-2727-7 [doi].

(33) Burrows NP. The molecular genetics of the Ehlers-Danlos syndrome. *Clin Exp Dermatol*. 1999;24(2):99-106.

(34) Byers PH, Duvic M, Atkinson M, et al. Ehlers-Danlos syndrome type VIIA and VIIB result from splice-junction mutations or genomic deletions that involve exon 6 in the COL1A1 and COL1A2 genes of type I collagen. *Am J Med Genet*. 1997;72(1):94-105.

(35) Canato M, Dal MM, Sbrana F, et al. Mechanical and electrophysiological properties of the sarcolemma of muscle fibers in two murine models of muscle dystrophy: col6a1-/- and mdx. *J Biomed Biotechnol*. 2010;2010:981945.

(36) Cartegni L, Wang J, Zhu Z, et al. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res*. 2003;31(13):3568-71.

(37) Chen HY, Chung YW, Lin WY, et al. Collagen type 3 alpha 1 polymorphism and risk of pelvic organ prolapse. *Int J Gynaecol Obstet*. 2008;103(1):55-8.

(38) Choi JK and Kim SC. Environmental effects on gene expression phenotype have regional biases in the human genome. *Genetics*. 2007;175(4):1607-13.

(39) Chou HT, Hung JS, Chen YT, et al. Association between COL3A1 collagen gene exon 31 polymorphism and risk of floppy mitral valve/mitral valve prolapse. *Int J Cardiol*. 2004;95(2-3):299-305.

(40) Collins M, Mokone GG, September AV, et al. The COL5A1 genotype is associated with range of motion measurements. *Scand J Med Sci Sports*. 2009;19(6):803-10.

(41) Collins M and Posthumus M. Type V collagen genotype and exercise-related phenotype relationships: a novel hypothesis. *Exerc Sport Sci Rev*. 2011;39(4):191-8.

(42) Collins M, Posthumus M and Schwellnus MP. The COL1A1 gene and acute soft tissue ruptures. *Br J Sports Med*. 2010;44(14):1063-4.

(43) Collins M, Xenophontos SL, Cariolou MA, et al. The ACE gene and endurance performance during the South African Ironman Triathlons. *Med Sci Sports Exerc*. 2004;36(8):1314-20.

(44) Corbin CB. Flexibility. *Clin J Sport Med*. 1984;3(1):101-17.

(45) Craib MW, Mitchell VA, Fields KB, et al. The association between flexibility and running economy in sub-elite male distance runners. *Med Sci Sports Exerc*. 1996;28(6):737-43.

(46) Decoster LC, Cleland J, Altieri C, et al. The effects of hamstring stretching on range of motion: a systematic literature review. *J Orthop Sports Phys Ther*. 2005;35(6):377-87.

(47) Doliana R, Mucignat MT, Segat D, et al. Alternative splicing of VWFA modules generates variants of type VI collagen alpha 3 chain with a distinctive expression pattern in embryonic chicken tissues and potentially different adhesive function. *Matrix Biol*. 1998;16(7):427-42.

(48) Dziadek M, Kazenwadel JS, Hendrey JA, et al. Alternative splicing of transcripts for the alpha 3 chain of mouse collagen VI: identification of an abundant isoform lacking domains N7-N10 in mouse and human. *Matrix Biol*. 2002;21(3):227-41.

(49) Eklund L, Piuhola J, Komulainen J, et al. Lack of type XV collagen causes a skeletal myopathy and cardiovascular defects in mice. *Proc Natl Acad Sci U S A*. 2001;98(3):1194-9.

(50) El Khoury L, Posthumus M, Collins M, et al. Polymorphic variation within the ADAMTS2, ADAMTS14, ADAMTS5, ADAM12 and TIMP2 genes and the risk of Achilles tendon pathology: A genetic association study. *J Sci Med Sport*. 2013;16(6):493-8.

(51) Erduran M, Altinisik J, Meric G, et al. Is Sp1 binding site polymorphism within COL1A1 gene associated with tennis elbow? *Gene*. 2014;537(2):308-11.

(52) Eriksen HA, Pajala A, Leppilahti J, et al. Increased content of type III collagen at the rupture site of human Achilles tendon. *J Orthop Res*. 2002;20(6):1352-7.

(53) Ficek K, Cieszczyk P, Kaczmarczyk M, et al. Gene variants within the COL1A1 gene are associated with reduced anterior cruciate ligament injury in professional soccer players. *J Sci Med Sport*. 2012;S1440-2440.

(54) Fitzgerald J and Bateman JF. A new FACIT of the collagen family: COL21A1. *FEBS Lett*. 2001;505(2):275-80.

(55) Fitzgerald J, Holden P and Hansen U. The expanded collagen VI family: new chains and new questions. *Connect Tissue Res*. 2013;54(6):345-50.

(56) Fleischmajer R, Perlish JS, Burgeson RE, et al. Type I and type III collagen interactions during fibrillogenesis. *Ann N Y Acad Sci*. 1990;580:161-75.

(57) Flynn RK, Pedersen CL, Birmingham TB, et al. The familial predisposition toward tearing the anterior cruciate ligament: a case control study. *Am J Sports Med*. 2005;33(1):23-8.

(58) Foster BP, Morse CI, Onambele GL, et al. Human COL5A1 rs12722 gene polymorphism and tendon properties in vivo in an asymptomatic population. *Eur J Appl Physiol*. 2014;114(7):1393-402.

(59) Frank CB. Ligament structure, physiology and function. *J Musculoskelet Neuronal Interact*. 2004;4(2):199-201.

(60) Gelse K, Poschl E and Aigner T. Collagens--structure, function, and biosynthesis. *Adv Drug Deliv Rev*. 2003;55(12):1531-46.

(61) Gianotti SM, Marshall SW, Hume PA, et al. Incidence of anterior cruciate ligament injury and other knee ligament injuries: a national population-based study. *J Sci Med Sport*. 2009;12(6):622-7.

(62) Gillies AR and Lieber RL. Structure and function of the skeletal muscle extracellular matrix. *Muscle Nerve*. 2011;44(3):318-31.

(63) Goeken LN and Hof AL. Instrumental straight-leg raising: results in healthy subjects. *Arch Phys Med Rehabil*. 1993;74(2):194-203.

(64) Gordon MK and Hahn RA. Collagens. *Cell Tissue Res*. 2010;339(1):247-57.

(65) Gostynski A, Pasmooij AM, Del RM, et al. Pigmentation and melanocyte supply to the epidermis depend on type XVII collagen. *Exp Dermatol*. 2014;23(2):130-2.

(66) Grahame R. Joint hypermobility syndrome pain. *Curr Pain Headache Rep*. 2009;13(6):427-33.

(67) Griffin LY, Albohm MJ, Arendt EA, et al. Understanding and preventing noncontact anterior cruciate ligament injuries: a review of the Hunt Valley II meeting, January 2005. *Am J Sports Med*. 2006;34(9):1512-32.

(68) Hagg PM, Hagg PO, Peltonen S, et al. Location of type XV collagen in human tissues and its accumulation in the interstitial matrix of the fibrotic kidney. *Am J Pathol*. 1997;150(6):2075-86.

(69) Hakim AJ, Cherkas LF, Grahame R, et al. The genetic epidemiology of joint hypermobility: a population study of female twins. *Arthritis Rheum*. 2004;50(8):2640-4.

(70) Hakim AJ and Grahame R. A simple questionnaire to detect hypermobility: an adjunct to the assessment of patients with diffuse musculoskeletal pain. *Int J Clin Pract*. 2003;57(3):163-6.

(71) Hay M, Patricios J, Collins R, et al. Association of type XI collagen genes with chronic Achilles tendinopathy in independent populations from South Africa and Australia. *Br J Sports Med*. 2013;47(9):569-74.

(72) Heikkinen A, Tu H and Pihlajaniemi T. Collagen XIII: a type II transmembrane protein with relevance to musculoskeletal tissues, microvessels and inflammation. *Int J Biochem Cell Biol*. 2012;44(5):714-7.

(73) Hewett TE, Ford KR, Myer GD, et al. Gender differences in hip adduction motion and torque during a single-leg agility maneuver. *J Orthop Res*. 2006;24(3):416-21.

(74) Hjorten R, Hansen U, Underwood RA, et al. Type XXVII collagen at the transition of cartilage to bone during skeletogenesis. *Bone*. 2007;41(4):535-42.

(75) Hoffmann A and Gross G. Tendon and ligament engineering in the adult organism: mesenchymal stem cells and gene-therapeutic approaches. *Int Orthop*. 2007;31(6):791-7.

(76) Hopkins WG. A New View of Statistics. *Ref Type: Online Source*. 2006.

(77) Hunter GR, Katsoulis K, McCarthy JP, et al. Tendon length and joint flexibility are related to running economy. *Med Sci Sports Exerc*. 2011;43(8):1492-9.

(78) Izu Y, Ansorge HL, Zhang G, et al. Dysfunctional tendon collagen fibrillogenesis in collagen VI null mice. *Matrix Biol*. 2010;30(1):53-61.

(79) Jarvinen TA, Jarvinen TL, Kaariainen M, et al. Muscle injuries: biology and treatment. *Am J Sports Med*. 2005;33(5):745-64.

(80) Jin H, Evangelou E, Ioannidis JP, et al. Polymorphisms in the 5' flank of COL1A1 gene and osteoporosis: meta-analysis of published studies. *Osteoporos Int*. 2011;22(3):911-21.

(81) Jin H, van't Hof RJ, Albagha OM, et al. Promoter and intron 1 polymorphisms of COL1A1 interact to regulate transcription and susceptibility to osteoporosis. *Hum Mol Genet*. 2009;18(15):2729-38.

(82) Jones AM. Running economy is negatively related to sit-and-reach test performance in international-standard distance runners. *International Journal of Sports Medicine*. 2002;23(1):40-3.

(83) Jozsa L and Kannus P. Human Tendons: Anatomy, Physiology and Pathology. *Human Kinetics, Champaign, IL*. 1997.

(84) Kadler KE, Baldock C, Bella J, et al. Collagens at a glance. *J Cell Sci*. 2007;120(Pt 12):1955-8.

(85) Kania AM, Reichenberger E, Baur ST, et al. Structural variation of type XII collagen at its carboxyl-terminal NC1 domain generated by tissue-specific alternative splicing. *J Biol Chem*. 1999;274(31):22053-9.

(86) Kannu P, Bateman JF, Randle S, et al. Premature arthritis is a distinct type II collagen phenotype. *Arthritis Rheum*. 2010;62(5):1421-30.

(87) Kannus P and Natri A. Etiology and pathophysiology of tendon ruptures in sports. *Scand J Med Sci Sports*. 1997;7(2):107-12.

(88) Kantarowski PG, Hiller WDB, Garret WE, et al. Cramping studies in 2600 endurance athletes. *Med Sci Sports Exerc*. 1990;22:S104.

(89) Kassner A, Hansen U, Miosge N, et al. Discrete integration of collagen XVI into tissue-specific collagen fibrils or beaded microfibrils. *Matrix Biol*. 2003;22(2):131-43.

(90) Kell RT, Bell G and Quinney A. Musculoskeletal fitness, health outcomes and quality of life. *Sports Med*. 2001;31(12):863-73.

(91) Khoschnau S, Melhus H, Jacobson A, et al. Type I collagen alpha1 Sp1 polymorphism and the risk of cruciate ligament ruptures or shoulder dislocations. *Am J Sports Med*. 2008;36(12):2432-6.

(92) Kjaer M. Role of extracellular matrix in adaptation of tendon and skeletal muscle to mechanical loading. *Physiol Rev*. 2004;84(2):649-98.

(93) Kjaer M, Langberg H, Heinemeier K, et al. From mechanical loading to collagen synthesis, structural changes and function in human tendon. *Scand J Med Sci Sports*. 2009;19(4):500-10.

(94) Kluivers KB, Dijkstra JR, Hendriks JC, et al. COL3A1 2209G>A is a predictor of pelvic organ prolapse. *Int Urogynecol J Pelvic Floor Dysfunct*. 2009;20(9):1113-8.

(95) Knapik JJ, Bauman CL, Jones BH, et al. Preseason strength and flexibility imbalances associated with athletic injuries in female collegiate athletes. *Am J Sports Med*. 1991;19(1):76-81.

(96) Knapik JJ, Jones BH, Bauman CL, et al. Strength, flexibility and athletic injuries. *Sports Med*. 1992;14(5):277-88.

(97) Koch M, Foley JE, Hahn R, et al. alpha 1(Xx) collagen, a new member of the collagen subfamily, fibril-associated collagens with interrupted triple helices. *J Biol Chem*. 2001;276(25):23120-6.

(98) Koch M, Laub F, Zhou P, et al. Collagen XXIV, a vertebrate fibrillar collagen with structural features of invertebrate collagens: selective expression in developing cornea and bone. *J Biol Chem*. 2003;278(44):43236-44.

(99) Koch M, Schulze J, Hansen U, et al. A novel marker of tissue junctions, collagen XXII. *J Biol Chem*. 2004;279(21):22514-21.

(100) Koch M, Veit G, Stricker S, et al. Expression of type XXIII collagen mRNA and protein. *J Biol Chem*. 2006;281(30):21546-57.

(101) Kong Q, Ma X, Li F, et al. COL6A1 polymorphisms associated with ossification of the ligamentum flavum and ossification of the posterior longitudinal ligament. *Spine (Phila Pa 1976)*. 2007;32(25):2834-8.

(102) Kongsgaard M, Qvortrup K, Larsen J, et al. Fibril morphology and tendon mechanical properties in patellar tendinopathy: effects of heavy slow resistance training. *Am J Sports Med*. 2010;38(4):749-56.

(103) Krivickas LS and Feinberg JH. Lower extremity injuries in college athletes: relation between ligamentous laxity and lower extremity muscle tightness. *Arch Phys Med Rehabil*. 1996;77(11):1139-43.

(104) Kubo K, Yata H and Tsunoda N. Effect of gene polymorphisms on the mechanical properties of human tendon structures. *Springer Plus*. 2013;2(1):343.

(105) Kumar P, Henikoff S and Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073-81.

(106) Kuo HJ, Maslen CL, Keene DR, et al. Type VI collagen anchors endothelial basement membranes by interacting with type IV collagen. *J Biol Chem*. 1997;272(42):26522-9.

(107) Laguette MJ, Abrahams Y, Prince S, et al. Sequence variants within the 3'-UTR of the *COL5A1* gene alters mRNA stability: Implications for musculoskeletal soft tissue injuries. *Matrix Biol*. 2011;30(5-6):338-45.

(108) Lahiri DK and Nurnberger JI, Jr. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res*. 1991;19(19):5554.

(109) Lampe AK and Bushby KM. Collagen VI related muscle disorders. *J Med Genet*. 2005;42(9):673-85.

(110) Lapiere CM, Nusgens B and Pierard GE. Interaction between collagen type I and type III in conditioning bundles organization. *Connect Tissue Res*. 1977;5(1):21-9.

(111) Lin DY, Hu Y and Huang BE. Simple and efficient analysis of disease association with missing genotype data. *Am J Hum Genet*. 2008;82(2):444-52.

(112) Lin DY, Zeng D and Millikan R. Maximum likelihood estimation of haplotype effects and haplotype-environment interactions in association studies. *Genet Epidemiol*. 2005;29(4):299-312.

(113) Little J, Higgins JP, Ioannidis JP, et al. STrengthening the REporting of Genetic Association studies (STREGA)--an extension of the STROBE statement. *Eur J Clin Invest*. 2009;39(4):247-66.

(114) Liu SH, Yang RS, al-Shaikh R, et al. Collagen in tendon, ligament, and bone healing. A current review. *Clin Orthop Relat Res*. 1995;(318):265-78.

(115) Liu X, Wu H, Byrne M, et al. Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development. *Proceedings of the National Academy of Sciences of the United States of America*. 1997;94(5):1852-6.

(116) Livak KJ and Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001;25(4):402-8.

(117) Lo IK, Marchuk L, Hart DA, et al. Messenger ribonucleic acid levels in disrupted human anterior cruciate ligaments. *Clin Orthop Relat Res*. 2003;(407):249-58.

(118) Macarthur DG and North KN. Genes and human elite athletic performance. *Hum Genet*. 2005;116(5):331-9.

(119) Maes HH, Beunen GP, Vlietinck RF, et al. Inheritance of physical fitness in 10-yr-old twins and their parents. *Med Sci Sports Exerc*. 1996;28(12):1479-91.

(120) Maganaris C, Narici M and Maffulli N. Biomechanics of the Achilles tendon. *Disabil Rehabil*. 2008;30:1542-7.

(121) Malfait F, Coucke P, Symoens S, et al. The molecular basis of classic Ehlers-Danlos syndrome: a comprehensive study of biochemical and molecular findings in 48 unrelated patients. *Hum Mutat*. 2005;25(1):28-37.

(122) Malfait F and De PA. Molecular genetics in classic Ehlers-Danlos syndrome. *Am J Med Genet C Semin Med Genet*. 2005;139C(1):17-23.

(123) Malfait F, Hakim AJ, De PA, et al. The genetic basis of the joint hypermobility syndromes. *Rheumatology (Oxford)*. 2006;45(5):502-7.

(124) Malfait F, Wenstrup RJ and De PA. Clinical and genetic aspects of Ehlers-Danlos syndrome, classic type. *Genet Med*. 2010;12(10):597-605.

(125) Manjra SI, Schwellnus MP and Noakes TD. Risk factors for exercise associated muscle cramping (EAMC) in marathon runners. *Med Sci Sports Exerc*. 1996;28(5 suppl):S167.

(126) Mann V, Hobson EE, Li B, et al. A COL1A1 Sp1 binding site polymorphism predisposes to osteoporotic fracture by affecting bone density and quality. *J Clin Invest*. 2001;107(7):899-907.

(127) Mazumder B, Seshadri V and Fox PL. Translational control by the 3'-UTR: the ends specify the means. *Trends Biochem Sci*. 2003;28(2):91-8.

(128) Mazzone MF and McCue T. Common conditions of the achilles tendon. *Am Fam Physician*. 2002;65(9):1805-10.

(129) Minamitani T, Ikuta T, Saito Y, et al. Modulation of collagen fibrillogenesis by tenascin-X and type VI collagen. *Exp Cell Res*. 2004;298(1):305-15.

(130) Mokone GG, Gajjar M, September AV, et al. The guanine-thymine dinucleotide repeat polymorphism within the tenascin-C gene is associated with achilles tendon injuries. *Am J Sports Med*. 2005;33(7):16-21.

(131) Mokone GG, Schwellnus MP, Noakes TD, et al. The COL5A1 gene and Achilles tendon pathology. *Scand J Med Sci Sports*. 2006;16(1):19-26.

(132) Myers JC, Li D, Bageris A, et al. Biochemical and immunohistochemical characterization of human type XIX defines a novel class of basement membrane zone collagens. *Am J Pathol*. 1997;151(6):1729-40.

(133) Nishiyama T, McDonough AM, Bruns RR, et al. Type XII and XIV collagens mediate interactions between banded collagen fibers in vitro and may modulate extracellular matrix deformability. *J Biol Chem*. 1994;269(45):28193-9.

(134) Norris F, Gasteiger E and Charfield P. An electromyographic study of induced and spontaneous muscle cramps. *Electroencephalogr Clin Neurophysiol*. 1956;9:139-47.

(135) Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet*. 2004;74(4):765-9.

(136) O'Brien M. Structure and metabolism of tendons. *Scand J Med Sci Sports*. 1997;7(2):55-61.

(137) O'Connell K, Knight H, Ficek K, et al. Investigation of Collagen Gene Variants with Anterior Cruciate Ligament Rupture in Two Independent Populations. *European Journal of Sport Science*. 2014;Accepted.

(138) O'Connell K, Posthumus M and Collins M. COL6A1 gene and Ironman triathlon performance. *Int J Sports Med*. 2011;32(11):896-901.

(139) O'Connell K, Posthumus M, Schwellnus MP, et al. Collagen genes and exercise-associated muscle cramping. *Clin J Sport Med*. 2013;23(1):64-9.

(140) Okada M, Kawahara G, Noguchi S, et al. Primary collagen VI deficiency is the second most common congenital muscular dystrophy in Japan. *Neurology*. 2007;69(10):1035-42.

(141) Olsen BR. New insights into the function of collagens from genetic analysis. *Curr Opin Cell Biol*. 1995;7(5):720-7.

(142) OMIM. Online Mendelian Inheritance in Man, OMIM®. *Ref Type: Online Source*. 2013.

(143) Ottani V, Raspanti M and Ruggeri A. Collagen structure and functional implications. *Micron*. 2001;32(3):251-60.

(144) Pace JM, Corrado M, Missero C, et al. Identification, characterization and expression analysis of a new fibrillar collagen gene, COL27A1. *Matrix Biol*. 2003;22(1):3-14.

(145) Parry DA, Barnes GR and Craig AS. A comparison of the size distribution of collagen fibrils in connective tissues as a function of age and a possible relation between fibril size distribution and mechanical properties. *Proc R Soc Lond B Biol Sci*. 1978;203(1152):305-21.

(146) Perenger TV. What's wrong with Bonferroni adjustments. *Br Med J*. 1998;316:1236-8.

(147) Pollitt R, McMahon R, Nunn J, et al. Mutation analysis of COL1A1 and COL1A2 in patients diagnosed with osteogenesis imperfecta type I-IV. *Hum Mutat*. 2006;27(7):716.

(148) Pollock M, Gaessar G, Butcher J, et al. American College of Sports Medicine Position Stand. The recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness, and flexibility in healthy adults. *Med Sci Sports Exerc*. 1998;30(6):975-91.

(149) Ponsioen TL, van Luyn MJ, van der Worp RJ, et al. Collagen distribution in the human vitreoretinal interface. *Invest Ophthalmol Vis Sci*. 2008;49(9):4089-95.

(150) Posthumus M, Collins M, Van der Merwe L, et al. Matrix metalloproteinase genes on chromosome 11q22 and the risk of anterior cruciate ligament (ACL) rupture. *Scand J Med Sci Sports*. 2012;22(4):523-33.

(151) Posthumus M, Schwellnus MP and Collins M. The COL5A1 gene: a novel marker of endurance running performance. *Med Sci Sports Exerc*. 2011;43(4):584-9.

(152) Posthumus M, September AV, Keegan M, et al. Genetic risk factors for anterior cruciate ligament ruptures: COL1A1 gene variant. *Br J Sports Med*. 2009;43(5):352-6.

(153) Posthumus M, September AV, O'Cuinneagain D, et al. The COL5A1 Gene Is Associated With Increased Risk of Anterior Cruciate Ligament Ruptures in Female Participants. *Am J Sports Med*. 2009;37(11):2234-40.

(154) Posthumus M, September AV, O'Cuinneagain D, et al. The association between the COL12A1 gene and anterior cruciate ligament ruptures. *Br J Sports Med*. 2010;44(16):1160-5.

(155) Posthumus M, September AV, Schwellnus MP, et al. Investigation of the Sp1-binding site polymorphism within the COL1A1 gene in participants with Achilles tendon injuries and controls. *J Sci Med Sport*. 2009;12(1):184-9.

(156) Posthumus M, September AV, Schwellnus MP, et al. The COL5A1 gene and musculoskeletal soft-tissue injuries. *South African J Sports Med*. 2010;22(2):38-41.

(157) Quarrie KL, Handcock P, Toomey MJ, et al. The New Zealand rugby injury and performance project. IV. Anthropometric and physical performance comparisons between positional categories of senior A rugby players. *Br J Sports Med*. 1996;30(1):53-6.

(158) Quarrie KL, Handcock P, Waller AE, et al. The New Zealand rugby injury and performance project. III. Anthropometric and physical performance characteristics of players. *Br J Sports Med*. 1995;29(4):263-70.

(159) Raine EV, Dodd AW, Reynard LN, et al. Allelic expression analysis of the osteoarthritis susceptibility gene COL11A1 in human joint tissues. *BMC Musculoskelet Disord*. 2013;14:85.

(160) Raleigh SM, Posthumus M, O'Cuinneagain D, et al. The GDF5 gene and anterior cruciate ligament rupture. *Int J Sports Med*. 2013;34(4):364-7.

(161) Reed UC, Ferreira LG, Liu EC, et al. Ullrich congenital muscular dystrophy and Bethlem myopathy: clinical and genetic heterogeneity. *Arquivos de Neuro-psiquiatria*. 2005;63(3B):785-90.

(162) Rees SG, Dent CM and Caterson B. Metabolism of proteoglycans in tendon. *Scand J Med Sci Sports*. 2009;19(4):470-8.

(163) Ribbans WJ and Collins M. Pathology of the tendo Achillis: do our genes contribute? *Bone Joint J*. 2013;95-B(3):305-13.

(164) Ricard-Blum S and Ruggiero F. The collagen superfamily: from the extracellular matrix to the cell membrane. *Pathologie-Biologie (Paris)*. 2005;53(7):430-42.

(165) Riley G. The pathogenesis of tendinopathy. A molecular perspective. *Rheumatology (Oxford)*. 2004;43(2):131-42.

(166) Riley G. Chronic tendon pathology: molecular basis and therapeutic implications. *Expert Rev Mol Med*. 2005;7(5):1-25.

(167) Roberts SP, Trewartha G, Higgitt RJ, et al. The physical demands of elite English rugby union. *J Sports Sci*. 2008;26(8):825-33.

(168) Rombaut L, Malfait F, Cools A, et al. Musculoskeletal complaints, physical activity and health-related quality of life among patients with the Ehlers-Danlos syndrome hypermobility type. *Disabil Rehabil*. 2010;32(16):1339-45.

(169) Sato K, Yomogida K, Wada T, et al. Type XXVI collagen, a new member of the collagen family, is specifically expressed in the testis and ovary. *J Biol Chem*. 2002;277(40):37678-84.

(170) Saunders CJ. Extracellular matric gene sequence variant analyses and Achilles tendinopathy. *PhD Thesis, University of Cape Town, Cape Town, South Africa*. 2013.

(171) Saunders CJ, Van der Merwe L, Posthumus M, et al. Investigation of variants within the COL27A1 and TNC genes and Achilles tendinopathy in two populations. *J Orthop Res*. 2013;31(4):632-7.

(172) Saunders CJ, Xenophontos SL, Cariolou MA, et al. The bradykinin beta 2 receptor (BDKRB2) and endothelial nitric oxide synthase 3 (NOS3) genes and endurance performance during Ironman Triathlons. *Hum Mol Genet*. 2006;15(6):979-87.

(173) Schug J. Using TESS to predict transcription factor binding sites in DNA sequence. In: Baxevanis AD, editor. Current protocols in bioinformatics. J. Wiley and Sons; 2003.

(174) Schwellnus MP. Cause of exercise associated muscle cramps (EAMC)--altered neuromuscular control, dehydration or electrolyte depletion? *Br J Sports Med*. 2009;43(6):401-8.

(175) Schwellnus MP, Allie S, Derman W, et al. Increased running speed and pre-race muscle damage as risk factors for exercise-associated muscle cramps in a 56 km ultra-marathon: a prospective cohort study. *Br J Sports Med*. 2011;45(14):1132-6.

(176) Schwellnus MP, Derman EW and Noakes TD. Aetiology of skeletal muscle 'cramps' during exercise: a novel hypothesis. *J Sports Sci*. 1997;15(3):277-85.

(177) Schwellnus MP, Drew N and Collins M. Muscle cramping in athletes--risk factors, clinical assessment, and management. *Clin Sports Med*. 2008;27(1):183-94.

(178) Schwellnus MP, Drew N and Collins M. Increased running speed and previous cramps rather than dehydration or serum sodium changes predict exercise-associated muscle cramping: a prospective cohort study in 210 Ironman triathletes. *Br J Sports Med*. 2011;45(8):650-6.

(179) September AV, Cook J, Handley CJ, et al. Variants within the COL5A1 gene are associated with achilles tendinopathy in two populations. *Br J Sports Med*. 2009;43:357-65.

(180) September AV, Nell EM, O'Connell K, et al. A pathway-based approach investigating the genes encoding interleukin-1beta, interleukin-6 and the interleukin-1 receptor antagonist provides new insight into the genetic susceptibility of Achilles tendinopathy. *Br J Sports Med*. 2011;45(13):1040-7.

(181) September AV, Posthumus M, Van der Merwe L, et al. The COL12A1 and COL14A1 genes and Achilles tendon injuries. *Int J Sports Med*. 2008;29(3):257-63.

(182) September AV, Schwellnus MP and Collins M. Tendon and ligament injuries: the genetic component. *Br J Sports Med*. 2007;41(4):241-6.

(183) Shang G, Collins M and Schwellnus MP. Factors associated with a self-reported history of exercise-associated muscle cramps in ironman triathletes: a case-control study. *Clin J Sport Med*. 2011;21(3):204-10.

(184) Slauterbeck JR, Hickox JR, Beynnon B, et al. Anterior cruciate ligament biology and its relationship to injury forces. *Orthop Clin North Am*. 2006;37(4):585-91.

(185) Smart DJ, Hopkins WG and Gill ND. Differences and changes in the physical characteristics of professional and amateur rugby union players. *J Strength Cond Res*. 2013;27(11):3033-44.

(186) Smith LT, Schwarze U, Goldstein J, et al. Mutations in the COL3A1 gene result in the Ehlers-Danlos syndrome type IV and alterations in the size and distribution of the major collagen fibrils of the dermis. *J Invest Dermatol*. 1997;108(3):241-7.

(187) Smith PJ, Zhang C, Wang J, et al. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet*. 2006;15(16):2490-508.

(188) Spivey KA, Banyard J, Solis LM, et al. Collagen XXIII: a potential biomarker for the detection of primary and recurrent non-small cell lung cancer. *Cancer Epidemiol Biomarkers Prev*. 2010;19(5):1362-72.

(189) Sund M, Vaisanen T, Kaukinen S, et al. Distinct expression of type XIII collagen in neuronal structures and other tissues during mouse development. *Matrix Biol*. 2001;20(4):215-31.

(190) Symoens S, Renard M, Bonod-Bidaud C, et al. Identification of binding partners interacting with the alpha1-N-propeptide of type V collagen. *Biochem J*. 2011;433(2):371-81.

(191) Tanaka T, Ikari K, Furushima K, et al. Genomewide linkage and linkage disequilibrium analyses identify COL6A1, on chromosome 21, as the locus

for ossification of the posterior longitudinal ligament of the spine. *Am J Hum Genet*. 2003;73(4):812-22.

(192) Tanaka T, Wakabayashi T, Oizumi H, et al. CLAC-P/collagen type XXV is required for the intramuscular innervation of motoneurons during neuromuscular development. *J Neurosci*. 2014;34(4):1370-9.

(193) Thacker SB, Gilchrist J, Stroup DF, et al. The impact of stretching on sports injury risk: a systematic review of the literature. *Med Sci Sports Exerc*. 2004;36(3):371-8.

(194) Timpl R and Brown JC. Supramolecular assembly of basement membranes. *Bioessays*. 1996;18(2):123-32.

(195) Tono-Oka S, Tanase S, Miike T, et al. Transient expression of collagen type XIV during muscle development and its reappearance after denervation and degeneration. *J Histochem Cytochem*. 1996;44(8):907-18.

(196) Tringali C, Brivio I, Stucchi B, et al. Prevalence of a characteristic gene profile in high-level rhythmic gymnasts. *J Sports Sci*. 2014;32(14):1409-15.

(197) Tsukahara S, Miyazawa N, Akagawa H, et al. COL6A1, the candidate gene for ossification of the posterior longitudinal ligament, is associated with diffuse idiopathic skeletal hyperostosis in Japanese. *Spine (Phila Pa 1976)*. 2005;30(20):2321-4.

(198) Tucker R and Collins M. What makes champions? A review of the relative contribution of genes and training to sporting success. *Br J Sports Med*. 2012;46(8):555-61.

(199) Tully LJ, Murphy AM, Smith RK, et al. Polymorphisms in TNC and COL5A1 genes are associated with risk of superficial digital flexor tendinopathy in National Hunt Thoroughbred racehorses. *Equine Vet J*. 2013;46(3):289-93.

(200) Ushiki T. Collagen fibers, reticular fibers and elastic fibers. A comprehensive understanding from a morphological viewpoint. *Arch Histol Cytol*. 2002;65(2):109-26.

(201) van der Rest M and Garrone R. Collagen family of proteins. *FASEB Journal*. 1991;5(13):2814-23.

(202) Veit G, Kobbe B, Keene DR, et al. Collagen XXVIII, a novel von Willebrand factor A domain-containing protein with many imperfections in the collagenous domain. *J Biol Chem*. 2006;281(6):3494-504.

(203) von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med*. 2007;4(10):e296.

(204) Wang SS, Whitney SL, Burdett RG, et al. Lower extremity muscular flexibility in long distance runners. *J Orthop Sports Phys Ther*. 1993;17(2):102-7.

(205) Warman ML, Abbott M, Apte SS, et al. A type X collagen mutation causes Schmid metaphyseal chondrodysplasia. *Nat Genet*. 1993;5(1):79-82.

(206) Wenstrup RJ, Florer JB, Davidson JM, et al. Murine model of the Ehlers-Danlos syndrome. col5a1 haploinsufficiency disrupts collagen fibril assembly at multiple stages. *J Biol Chem*. 2006;281(18):12888-95.

(207) Wenstrup RJ, Smith SM, Florer JB, et al. Regulation of collagen fibril nucleation and initial fibril assembly involves coordinate interactions with collagens V and XI in developing tendon. *J Biol Chem*. 2011;286(23):20455-65.

(208) Witvrouw E, Danneels L, Asselman P, et al. Muscle flexibility as a risk factor for developing muscle injuries in male professional soccer players. A prospective study. *Am J Sports Med*. 2003;31(1):41-6.

(209) Wu JJ, Woods PE and Eyre DR. Identification of cross-linking sites in bovine cartilage type IX collagen reveals an antiparallel type II-type IX molecular relationship and type IX to type IX bonding. *J Biol Chem*. 1992;267(32):23007-14.

(210) Xie X, Lu J, Kulbokas EJ, et al. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*. 2005;434(7031):338-45.

(211) Ylikarppa R, Eklund L, Sormunen R, et al. Lack of type XVIII collagen results in anterior ocular defects. *FASEB J*. 2003;17(15):2257-9.

(212) Youdas JW, Krause DA and Hollman JH. Validity of hamstring muscle length assessment during the sit-and-reach test using an inclinometer to measure hip joint angle. *J Strength Cond Res*. 2008;22(1):303-9.

(213) Young BB, Zhang G, Koch M, et al. The roles of types XII and XIV collagen in fibrillogenesis and matrix assembly in the developing cornea. *J Cell Biochem*. 2002;87(2):208-20.

(214) Zazulak BT, Paterno M, Myer GD, et al. The effects of the menstrual cycle on anterior knee laxity: a systematic review. *Sports Med*. 2006;36(10):847-62.

(215) Zeng D, Lin DY, Avery CL, et al. Efficient semiparametric estimation of haplotype-disease associations in case-cohort and nested case-control studies. *Biostatistics*. 2006;7(3):486-502.

(216) Zhou Z and Fu XD. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*. 2013;122(3):191-207.

(217) Zoppi N, Gardella R, De PA, et al. Human fibroblasts with mutations in COL5A1 and COL3A1 genes do not organize collagens and fibronectin in the extracellular matrix, down-regulate alpha2beta1 integrin, and recruit alphavbeta3 Instead of alpha5beta1 integrin. *J Biol Chem*. 2004;279(18):18157-68.

(218) Zou Y, Zwolanek D, Izu Y, et al. Recessive and dominant mutations in COL12A1 cause a novel EDS/myopathy overlap syndrome in human and mice. *Hum Mol Genet*. 2013;23(9):2339-52.